

Nonlinear full waveform inversion of wide-aperture OBS data for Moho structure using a trans-dimensional Bayesian method

Peng Guo¹,¹ Satish C. Singh,² Venkata A. Vaddineni,² Gerhard Visser¹,¹
Ingo Grevenmeyer³ and Erdinc Saygin^{1,4}

¹Deep Earth Imaging Future Science Platform, The Commonwealth Scientific and Industrial Research Organisation (CSIRO), Kensington 6151, Australia.

E-mail: peng.guo@csiro.au

²Laboratoire de Géosciences Marines, Institut de Physique du Globe de Paris, Paris 75005, France

³GEOMAR Helmholtz Centre for Ocean Research, D-24148 Kiel, Germany

⁴Department of Physics, School of Physics, Mathematics and Computing, Faculty of Engineering and Mathematical Sciences, University of Western Australia, Crawley 6009, Australia

Accepted 2020 October 14. Received 2020 October 8; in original form 2020 June 6

SUMMARY

Seismic full waveform inversion (FWI) is a powerful method for estimating quantitative subsurface physical parameters from seismic data. As the FWI is a nonlinear problem, the linearized approach updates model iteratively from an initial model, which can get trapped in local minima. In the presence of a high-velocity contrast, such as at Moho, the reflection coefficient and recorded waveforms from wide-aperture seismic acquisition are extremely nonlinear around critical angles. The problem at the Moho is further complicated by the interference of lower crustal (Pg) and upper mantle (Pn) turning ray arrivals with the critically reflected Moho arrivals (PmP). In order to determine velocity structure near Moho, a nonlinear method should be used. We propose to solve this strong nonlinear FWI problem at Moho using a trans-dimensional Markov chain Monte Carlo (MCMC) method, where the earth model between lower crust and upper mantle is ideally parametrized with a 1-D assumption using a variable number of velocity interfaces. Different from common MCMC methods that require determining the number of unknown as a fixed prior before inversion, trans-dimensional MCMC allows the flexibility for an automatic estimation of both the model complexity (e.g. the number of velocity interfaces) and the velocity–depth structure from the data. We first test the algorithm on synthetic data using four representative Moho models and then apply to an ocean bottom seismometer (OBS) data from the Mid-Atlantic Ocean. A 2-D finite-difference solution of an acoustic wave equation is used for data simulation at each iteration of MCMC search, for taking into account the lateral heterogeneities in the upper crust, which is constrained from traveltimes tomography and is kept unchanged during inversion; the 1-D model parametrization near Moho enables an efficient search of the trans-dimensional model space. Inversion results indicate that, with very little prior and the wide-aperture seismograms, the trans-dimensional FWI method is able to infer the posterior distribution of both the number of velocity interfaces and the velocity–depth model for a strong nonlinear problem, making the inversion a complete data-driven process. The distribution of interface matches the velocity discontinuities. We find that the Moho in the study area is a transition zone of 0.7 km, or a sharp boundary with velocities from around 7 km s^{−1} in the lower crust to 8 km s^{−1} of the upper mantle; both provide nearly identical waveform match for the field data. The ambiguity comes from the resolution limit of the band-limited seismic data and limited offset range for PmP arrivals.

Key words: Probability distributions; Waveform inversion; Body waves; Crustal imaging.

1 INTRODUCTION

Seismic full waveform inversion (FWI) is a state-of-the-art method for estimating high-resolution quantitative subsurface images by fully exploiting the recorded waveform (Tarantola 1984; Pratt *et al.* 1998; Shipp & Singh 2002; Fichtner *et al.* 2006; Virieux & Operto 2009; Ray *et al.* 2017). It is based on minimizing the difference between observed and synthetic data, with a numerical solution of the complete wave equation for realistic simulation of seismic wave propagation. A gradient-based linearized iterative inversion approach is widely used, where the adjoint method is applied to calculate the gradients of the data misfit by cross-correlation of the forward and adjoint wavefields (Tromp *et al.* 2005; Fichtner *et al.* 2006; Plessix 2006). The linearized approach has been used for first arrival turning rays (refraction) and pre-critical reflection data (Shipp & Singh 2002; Zelt *et al.* 2003; Vigh *et al.* 2014). However, with limited data coverage and frequency bandwidth, FWI is a highly nonlinear and non-unique problem (Bozdağ *et al.* 2011). Previous studies suggest that the final outcome from FWI under linear assumption highly depends on the starting model, and the iterative process can get trapped easily in a local minima (Bozdağ *et al.* 2011; Métivier *et al.* 2016).

In the presence of a high-velocity contrast, such as at Moho, the reflection coefficient and recorded waveforms from wide-aperture seismic acquisition are extremely nonlinear. The elastic reflection coefficient becomes complex starting from the critical angle, with dramatic changes in both the amplitude and phase of reflected waves at near- and post-critical incident angles, which do not lend themselves to linearization (Aki & Richards 1980). Significant changes in the waveform shape can occur at critical angles, where the waveform of a reflection will be the corresponding Hilbert-transform of the incident wave. The problem at the Moho is further complicated by the interference of lower crustal (Pg) and upper mantle (Pn) turning ray arrivals with the critically reflected Moho arrivals (PmP; McGlashan *et al.* 2008; Buehler & Shearer 2017). In order to determine velocity structure near Moho, a nonlinear method should be used.

Bayesian inference using Markov chain Monte Carlo (MCMC) is a fully nonlinear inversion method based on an extensive search of the model space, which is able to remove the dependence on the starting model, and also provide uncertainty analysis for the inversion results. The feasibility of seismic inversion using MCMC sampling has been demonstrated by several authors (Cary & Chapman 1988; Mosegaard & Tarantola 1995; Sambridge & Mosegaard 2002; Fichtner & Zunino 2019), with a recently growing interest of using advanced MCMC methods for applications in geoscience, including, for example, trans-dimensional (reversible-jump) MCMC (Green 1995; Malinverno 2002; Bodin & Sambridge 2009; Sambridge *et al.* 2013; Hawkins & Sambridge 2015; Saygin *et al.* 2015; Dettmer *et al.* 2016; Galetti *et al.* 2017; Ray *et al.* 2017; Guo *et al.* 2020) and Hamiltonian MCMC (Sen & Biswas 2017; Fichtner *et al.* 2018; Gebraad *et al.* 2020).

With regard to velocity model building near Moho using Bayesian methods, Cary & Chapman (1988) used MCMC sampling for inverting Moho models and associated error analysis from seismic refraction waveforms, where the parameters being estimated were the depths of a number of fixed velocities. The fixed velocities and the number of layers limit the flexibility of the inversion algorithm, making the search space affected by subjective human intervention. Mosegaard *et al.* (1997) applied the MCMC technique for Moho structure inversion using near-normal-incidence reflections, where the problem is only moderately nonlinear. Besides MCMC

sampling, other nonlinear global optimization methods have also been used for addressing the strong nonlinearity observed in the Moho model inversion. Sambridge & Drijkoningen (1992) applied a genetic algorithm for 1-D Moho model inversion from marine refraction data, and showed better convergence rate than the classic MCMC method.

Inversion methods require solving the forward modelling problem for computing the predicted seismic data, for a given earth model at each iteration. Nonlinear inversion using MCMC sampling usually requires a large number of iterations for convergence. Limited by computational resources, approximations for wave equations are usually used in previous studies for nonlinear seismic inversion, especially when targeted for deep earth model building such as Moho. Mosegaard *et al.* (1997) used a propagator matrix method (Ganley 1981) for predicting seismograms at each iteration of MCMC sampling. The WBKJ method (Chapman 1978; Chapman & Orcutt 1985) was used for efficiently calculating synthetic seismograms in both Cary & Chapman (1988) and Sambridge & Drijkoningen (1992). Cary & Chapman (1988) compared the accuracy of the WKBJ method with synthetic seismograms calculated from the reflectivity method, and indicated that the it was accurate enough for the data used in their study. The solution from WKBJ is, nonetheless, still an approximation to the full solution of wave equation, therefore inaccuracy may appear in the presence of strong velocity variations (Shaw 1986), such as the Moho boundary, which causes the systematically smoothed models of high-velocity gradient from inversion as discussed in Cary & Chapman (1988). Therefore, a more accurate solution for forward modelling should be used for the nonlinear Moho FWI problem, when the capability of high-performance computing (HPC) allows.

The fast development of HPC capabilities enables the emergence and accelerated application of advanced model space search algorithms for inversion, and using more accurate forward modelling methods for better description of the wave physics. In the present study, we use a trans-dimensional MCMC method for inferring the posterior distribution of earth models near Moho, which is idealized with a 1-D model assumption for simplicity and illustration, from mainly PmP phase around critical angles in the wide-aperture seismic data that also contain Pg and Pn arrivals. Besides the velocity–depth structure of the earth model, model dimension (the number of velocity interfaces) is also a variable to be inferred from data, which makes the outcome from trans-dimensional FWI a more complete solution for the inversion problem. At each iteration of the MCMC chain, a time-domain finite-difference solution of 2-D wave equation is used for predicting seismograms, for taking into account the relatively strong horizontal heterogeneities in the upper crust. We provide synthetic studies using four representative Moho structure models, followed by a field data study using a long-offset (approximately 40 km) OBS seismic data from the Mid-Atlantic Ocean for inverting Moho structures for crust formed at slow spreading Mid-Atlantic Ridge (MAR).

2 METHODOLOGY

In this section, we formulate the FWI problem in a Bayesian framework, where all the information involved is represented using probabilistic terms (Gouveia & Scales 1998). In contrast to obtaining a single final physical model of the subsurface using common linearized inversion methods, from the observed seismic data **d**, we aim to estimate the posterior distribution of the seismic velocity model **m**, from which we can derive a probabilistic description of

the earth model, including for example, the ‘optimal’ solution and uncertainty quantification.

Based on the Bayes’ theorem, the posterior distribution can be defined as

$$p(\mathbf{m}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{m})p(\mathbf{m})}{p(\mathbf{d})}, \quad (1)$$

where $p(\mathbf{m})$ is the prior distribution of the model parameters, which represents the knowledge about the model before measuring the data. $p(\mathbf{d}|\mathbf{m})$ is the likelihood function, which is the probability of observed data \mathbf{d} with simulated data given model \mathbf{m} . Thus, the posterior distribution reflects how the prior knowledge about the model parameters is updated by the observed data, which, in the context of FWI, is the recorded waveform information. $p(\mathbf{d})$ is a normalizing constant also known as evidence, that makes the integral of the posterior distribution equal to unity. The computation of $p(\mathbf{d})$ is usually prohibitive, as it requires integration over the model space.

The most common technique for Bayesian inference is probably MCMC. Because it only requires computation of the probability ratio for pairs of models, and $p(\mathbf{d})$ is independent of a specific model, we can neglect the evidence term and replace eq. (1) with

$$p(\mathbf{m}|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{m})p(\mathbf{m}). \quad (2)$$

2.1 Model parametrization

MCMC requires a large number of iterations for sampling the posterior distribution, and the required iteration can increase dramatically in an exponential scale with the increasing number of parameters [model dimension; the curse of dimensionality (Sambridge *et al.* 2006)]. Seismic inversion is typically formulated as a high-dimensional problem, therefore a sparse model parametrization is required for an efficient search of the model space.

The velocity structure near Moho (the lower crust, Moho and the upper mantle) can be approximately considered as horizontally invariant within the offset range of a common OBS gather for OBS data. Therefore we use a 1-D model assumption defined by a number of velocity nodes (interface discontinuities); these velocity nodes are used for interpolating velocity values at depths between two neighbouring nodes. The model in depth between two neighbouring nodes can be considered as a layer.

We use cubic spline interpolation for deriving 1-D velocity models from the sparsely distributed velocity nodes for simplicity (Fig. 1a). Cubic spline interpolation is able to describe smoothing velocity transitions, as is the case for many models of the subsurface, but may have difficulties when there is an abrupt (e.g. a step function) velocity change, such as at Moho. Thus we introduce a second model parametrization, that the cubic spline interpolation is applied for the shallower part, and from certain depth (9 km in this study) of the lower crust to the upper mantle, a step function is used for interpolating velocities between two neighbouring velocity nodes. With a step function, the velocities between two neighbouring velocity nodes are equal to the value of the node at greater depth (Fig. 1b). We refer to the second approach for defining velocity models as a ‘hybrid’ (cubic spline interpolation for shallower part and a step function for greater depth) model parametrization.

Besides inverting for the properties (including depth and velocity) of velocity nodes from the observed seismic data, the trans-dimensional MCMC method is also able to infer the number of velocity nodes (model dimensions) required for describing the velocity structure complexity, as indicated by the name ‘trans-dimensional’.

We will introduce the specific Bayesian method in the following section.

2.2 Trans-dimensional MCMC

Given observed data, MCMC sampling is an iterative process to generate model samples from a target distribution, where each of the models is a perturbation of the previous one. We use trans-dimensional MCMC sampling to explore the model space (Green 1995; Malinverno 2002; Agostinetti & Malinverno 2010; Burdick & Lekić 2017; Killingbeck *et al.* 2018). The attractive property of this particular sampling method is that, besides physical properties of velocity models (velocity–depth structure for 1-D model), the model dimension (the number of velocity nodes for our case) is also a variable to be inferred from data; thus the algorithm greatly simplifies the preparation work for determining such parameters (Ray *et al.* 2016; Dadi *et al.* 2018; Hawkins *et al.* 2019). Trans-dimensional MCMC can be considered as an extension of the popular Metropolis–Hastings algorithm, with proposal functions that can increase or decrease model dimension (Bodin *et al.* 2012). Compared with conventional MCMC method for seismic inversion that requires determining the number of unknowns (e.g. the number of layers for 1-D model) before inversion starts (Mosegaard *et al.* 1997), trans-dimensional MCMC provides a more complete data-driven solution for the FWI problem, with outcomes including posterior distribution for both the model complexity and physical properties.

In the context of 1-D velocity model inversion, we list the four general types of model update proposals for creating new models in the implementation of the trans-dimensional MCMC as

- (i) Birth: a new velocity node is inserted at a random depth (increasing of model dimension), and therefore a randomly selected layer is divided into two.
- (ii) Death: a randomly selected velocity node is removed (decreasing of model dimension), and therefore two neighbouring layers are merged together.
- (iii) Node depth: depth of a randomly selected velocity node is adjusted to shallower or deeper depth, using perturbation values from a Gaussian distribution.
- (iv) Seismic velocity: the velocity of a randomly selected velocity node is adjusted, with perturbation values from a Gaussian distribution.

The sampled model at each iteration of MCMC sampling is a series of velocity nodes with different velocity values, which are distributed at different depths. As defined before, a node represents a velocity discontinuity, and the velocity between two neighbouring nodes can be defined using interpolation; the 1-D model between two neighbouring nodes can be viewed as a layer. The interpolated velocity–depth model from the sparsely distributed velocity nodes is used for seismic modelling for predicting data.

For trans-dimensional FWI, the starting velocity model of a chain is selected randomly from the prior distribution. At each iteration of an MCMC chain, a new model is proposed by perturbing the current state of model parameters, using a proposal function listed above. The new model will either be rejected, or accepted as the latest member of the Markov chain depending on the acceptance criterion ratio (Green 1995; Ray *et al.* 2017; Hawkins *et al.* 2019),

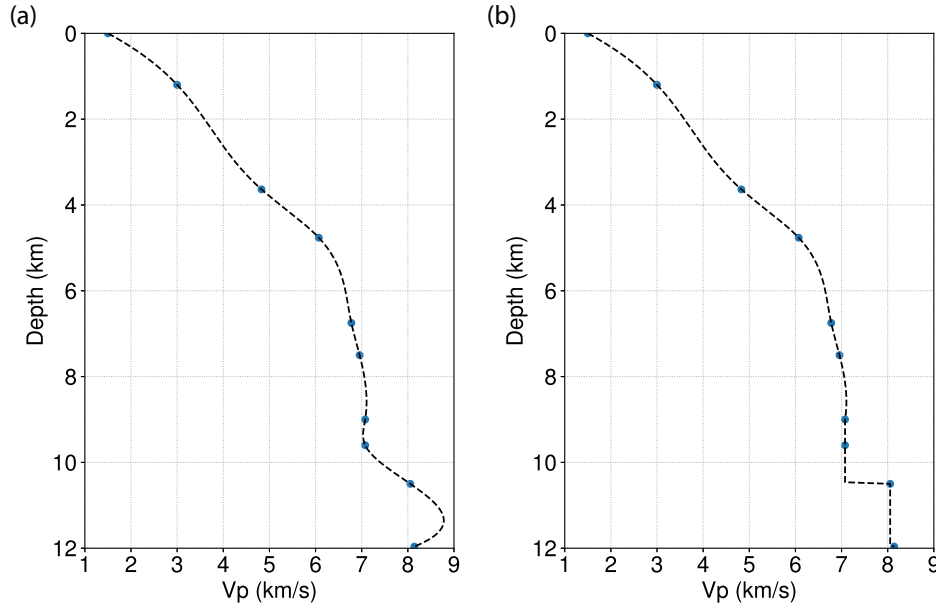


Figure 1. The 1-D velocity model can be defined using a series of sparsely distributed velocity nodes, with (a) the cubic spline interpolation and (b) the hybrid model parametrization.

as

$$\begin{aligned} \alpha(\mathbf{m}'|\mathbf{m}) &= \min [1, \text{Prior ratio} \times \text{Likelihood ratio} \\ &\quad \times \text{Proposal ratio} \times |\mathbf{J}|] \\ &= \min \left[1, \frac{p(\mathbf{m}')}{p(\mathbf{m})} \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})} \frac{Q(\mathbf{m}' \rightarrow \mathbf{m})}{Q(\mathbf{m} \rightarrow \mathbf{m}')} \times |\mathbf{J}| \right], \end{aligned} \quad (3)$$

where the term $\frac{p(\mathbf{m}')}{p(\mathbf{m})}$ is the prior ratio, $\frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})}$ the likelihood ratio and $\frac{Q(\mathbf{m}' \rightarrow \mathbf{m})}{Q(\mathbf{m} \rightarrow \mathbf{m}')}$ the proposal ratio. The proposal ratio measures the probability that perturbs the current model \mathbf{m} to obtain the new proposed model \mathbf{m}' , calculated by the posterior distribution evaluated at \mathbf{m}' to \mathbf{m} multiplied by the ratio for the reverse step from \mathbf{m} to \mathbf{m}' (Sambridge *et al.* 2006; Dettmer *et al.* 2010; Zhu & Gibson 2018). The difference in the acceptance ratio with the common Metropolis–Hastings algorithm is the extra term $|\mathbf{J}|$, the determinant of the Jacobian when transformation from \mathbf{m} to \mathbf{m}' , which is required to preserve volume when the transformation involves a jump between dimensions (Green 2003; Bodin & Sambridge 2009). When a Markov chain is completed, the first part of the chain, that is, the ‘burn-in’ stage, will be discarded; after that, the chain is assumed to be in the stationary phase, where the model samples will be considered from the posterior distribution constrained by the prior and data. The model ensemble from MCMC FWI contains velocity models from the posterior distribution, from which the probabilistic properties for the inversion problem can be derived.

2.3 Forward problem

Inversion methods require solving the forward modelling problem of computing the predicted data for an earth model at each iteration. The acoustic assumption has been widely applied for simulating seismic wave propagation in FWI problems (Tarantola 1984; Virieux & Operto 2009; Kamei *et al.* 2013; Ray *et al.* 2017). We use the first-order derivative formulation with pressure and particle

velocities (Auld 1973) as

$$\begin{aligned} \frac{\partial p}{\partial t} &= -\rho c_p^2 (\nabla \cdot \mathbf{v}), \\ \frac{\partial \mathbf{v}}{\partial t} &= -\frac{1}{\rho} \nabla p, \end{aligned} \quad (4)$$

where p is the pressure wavefield, \mathbf{v} is the particle velocity vector, ρ is density, and c_p is the P wave velocity. In this study, the time-domain staggered-grid finite-difference (FD; Virieux 1986) is used for solving the acoustic wave equations for predicting seismic data given a model, with eighth-order accuracy in space, and second-order accuracy in time. The forward modelling solver must be numerically stable for all the models allowed by the prior, therefore a conservative choice of temporal and spatial size is chosen for satisfying the stability condition of finite-difference simulation. The convolutional perfectly matched layer (CPML) (Martin & Komatitsch 2009) is used as the absorbing boundary condition. Note that while 1-D modelling is much more computationally efficient, here we use 2-D FD modelling for a better description of the wave physics in the oceanic crust with pronounced horizontal heterogeneities.

2.4 Likelihood function

The likelihood function is based on a least-squares measure of the data misfit, defined as $\mathbf{r} = \mathbf{d} - f(\mathbf{m})$, where \mathbf{d} is the observed pre-stack seismic shot gathers, and $f(\mathbf{m})$ is the simulated seismic data with model \mathbf{m} . The data errors are approximated by a covariance matrix \mathbf{C} . The Gaussian likelihood function is

$$p(\mathbf{d}|\mathbf{m}) = \frac{1}{\sqrt{(2\pi)^{N_s} |\mathbf{C}|}} \exp \left(-\frac{1}{2} \mathbf{r}^T \mathbf{C}^{-1} \mathbf{r} \right), \quad (5)$$

where N_s is the number of samples in the seismic data. The covariance matrix \mathbf{C} is usually assumed to be diagonal for seismic FWI. However, the noise is temporally correlated in seismic records, thus \mathbf{C} is not diagonal (Sambridge 1999; Agostinetti & Malinverno

2010). We use a 1-D Gaussian process for describing the temporally correlated noise, where the element of \mathbf{C} is parametrized with a stationary Gaussian variogram (Bodin *et al.* 2012; Visser *et al.* 2019).

2.5 The prior

The prior contains information we know about the model parameters before measuring data, and defines the model space to be searched by Markov chains; for seismic inversion, it is usually determined with information from previous work, and knowledge of the expected subsurface physical parameter ranges.

In this study, we use prior distributions controlled by mean and standard deviation parameters, which are able to provide flexibility for defining distributions over models, while at the same time keeping the number of required prior parameters small (Visser *et al.* 2019).

We use independent Gamma distributions as priors over each node's velocity v_i , with the probability for v defined as

$$p(v; \alpha, \beta) = \frac{\beta^\alpha v^{\alpha-1} e^{-\beta v}}{\Gamma(\alpha)}, \quad (6)$$

the mean is defined $\mu = \alpha/\beta$ and a standard deviation $\sigma = \mu/\sqrt{\alpha}$; the Gamma distribution has the benefit of allowing only real positive numbers. Additional constraints, including the minimum and maximum velocities, are used to avoid sampling unrealistic velocity values.

The vector of layer width \mathbf{w} (the distance in depth between two neighbouring velocity nodes) has a symmetric Dirichlet prior distribution,

$$p(\mathbf{w}|\mathbf{k}) = \frac{1}{B(\mathbf{v}, \mathbf{k}) L^k} \prod_{i=1}^k \left(\frac{w_i}{L}\right)^{v_i-1} \quad (7)$$

where w is the distance in depth between two neighbouring velocity nodes (the layer width); the vector of layer width contains k values, with their sum equal to the maximum model depth L for inversion. $B(\mathbf{v}, \mathbf{k})$ is a Beta function serving as a normalizing constant. v is a parameter defining the shape of the prior density; we set $v = 2$ for simplicity.

A Poisson distribution is used as the prior for the model dimension k (the number of velocity nodes / interfaces) as

$$p(k) = e^{-\lambda} \frac{\lambda^{k-1}}{(k-1)!} \quad (8)$$

where λ is the expected number of velocity nodes. We also set a maximum value of N .

3 SYNTHETIC EXAMPLES

The seismic discontinuity at Moho is a primary boundary between the crust and upper mantle, defined mainly by seismic wave velocities. In general, the P wave seismic velocity in the lower crust is about 7 km s^{-1} , and in the upper mantle around 8 km s^{-1} ; therefore, a large velocity contrast at Moho up to $1\text{--}1.5 \text{ km s}^{-1}$ is expected (Grad *et al.* 2009). We use four 1-D velocity profiles, shown in Fig. 2, for synthetic studies. These velocity models represent four different scenarios for the Moho structure, with (a) a sharp boundary, (b) a smooth transition, (c) step functions with increasing velocity and (d) with low-velocity anomalies in the upper mantle. A 1-D velocity profile is extended to a 2-D horizontally invariant velocity model for

imaging with 2-D seismic FWI using far-offset seismic data with critically reflected waves and upper-mantle refractions.

Fig. 3 shows the acquisition geometry for generating seismic data in synthetic studies. The ocean bottom seismometer (OBS) is located at 3.25 km depth and 0.5 km model distance, on the seafloor of a sedimentary layer. The water depth is 3.25 km, and the oceanic crust is about 6 km thick. There are 120 sources, evenly distributed from offsets 12 to 36 km. We use this offset range because best quality of OBS signal are observed in this offset range for critically reflected PmP arrivals. For the purpose of seismic modelling, the OBS, which is used for recording seismic waves, is considered as the 'source' using reciprocity, and the sources near to the water surface are treated as 'receivers'. The offset corresponding to the critical angle for the Moho reflection is around 28 km. For inversion, the velocities shallower than 6 km in depth (about 2 km from ocean bottom) are set to the correct value because the relatively far offset data we used has fewer constraints for the upper crust; we only invert for those of the deeper (from 6 to 12 km) part of the model, including the lower crust, Moho and the upper mantle.

Two kinds of model parametrizations are used for interpolating velocity values at finite-difference grids from the sparsely distributed velocity nodes from MCMC search, i.e., the cubic spline interpolation method and the hybrid [cubic spline interpolation (6–9 km depth) and step function (9–12 km depth)] method, as described in the previous section. For all the four examples, we use 20 independent trans-dimensional MCMC chains, each with 100 000 iterations. The MCMC sampler starts from a constant velocity model of 6.6 km s^{-1} , which is very far away from the true solution. The first 60 000 iterations are considered as the 'burn-in' stage and the sampled models are discarded; only the velocity models from the final 40 000 iterations of each chain are included in the recorded ensemble of models.

The prior for the number of velocity nodes is a Poisson distribution, with 10 as the expected number and a maximum value of 20. The vector of layer width (the distance in depth between two neighbouring velocity nodes) is defined with a Dirichlet prior distribution, with the sum equal to 6 km. Prior for the velocity values are bounded between 6 and 8.6 km s^{-1} with a mean of 7 km s^{-1} and a standard deviation of 1 km s^{-1} for a Gamma distribution for all the depths.

3.1 Example 1

Fig. 4(b) shows an OBS gather generated using the velocity model of Moho structure with a sharp boundary (in Fig. 4a, the same with Fig. 2a), and Fig. 4(d) shows the predicted OBS gather using the starting model (Fig. 4c) for the MCMC sampler; Moho reflection (PmP) and refraction (Pn) is missing in the data from the starting model. The seismic data (Fig. 4b) from the true model added with random temporally correlated noise (Bodin *et al.* 2012) is used as the 'observed' data for inversion, with a comparison of a few selected seismic traces with and without adding noise shown in Fig. 5. The signal-to-noise ratio (SNR; calculated as the ratio between the maximum PmP amplitude and the root mean square in a 1.0 s time window before the PmP arrival) is 4.8.

Velocity–depth profiles are interpolated from sparsely distributed velocity nodes, where the required number of velocity nodes for describing the model complexity and the depth and velocity value of each node are inferred from trans-dimensional MCMC. Fig. 6 shows the posterior probability density of the interpolated velocity models using velocity nodes from trans-dimensional MCMC, with the cubic

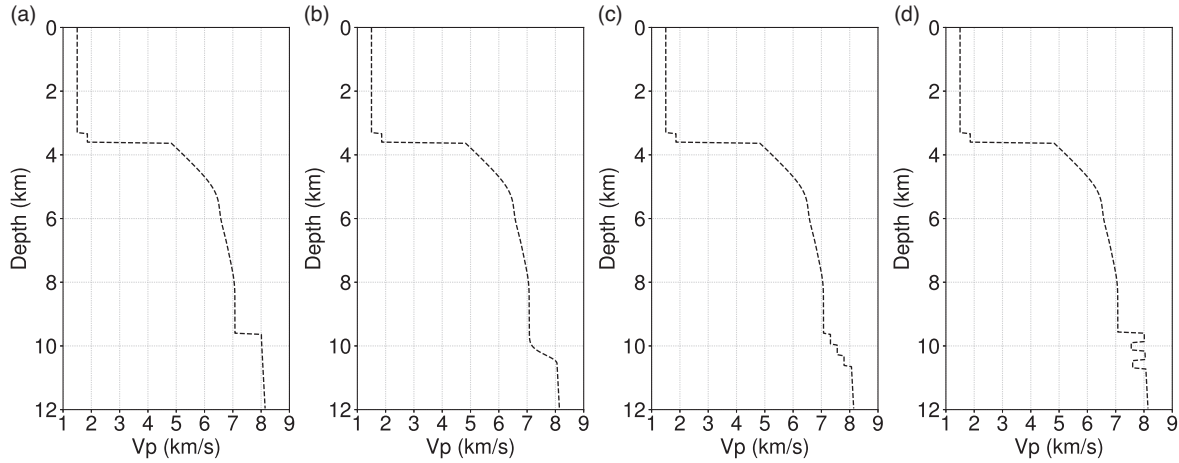


Figure 2. 1-D velocity models representing four different Moho structures for synthetic tests, with (a) a sharp boundary, (b) a smooth transition, (c) step functions with increasing velocity and (d) with low-velocity anomalies in the upper mantle.

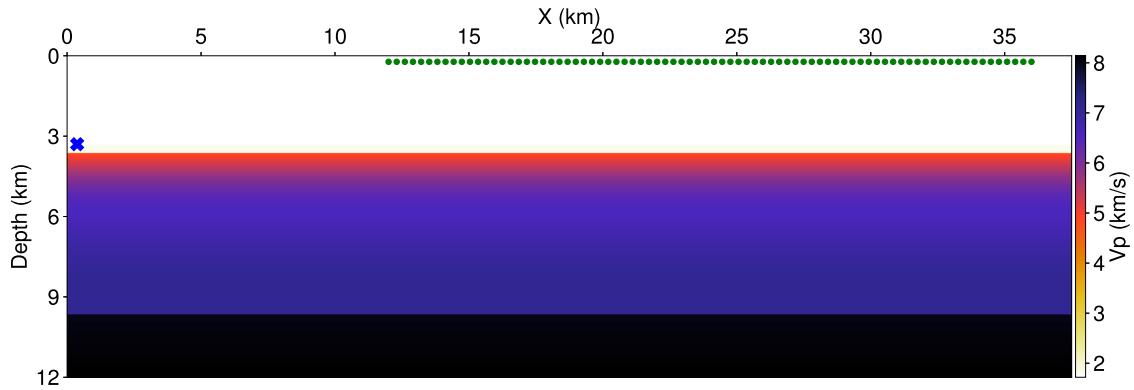


Figure 3. 2-D model with horizontally invariant velocities, using the velocity profile in Fig. 2(a). The blue symbol refers to location of the ocean bottom seismometer (OBS) for recording seismic waves, and the green dots near the water surface refer to seismic sources. Using reciprocity, the OBS is treated as the ‘source’, and the sources close to the water surface are treated as ‘receivers’ for seismic modelling.

spline interpolation (Fig. 6a) and the hybrid model parametrization (Fig. 6c), respectively. Following Mosegaard *et al.* (1997), we retain models every 100 iterations between accepted models for estimating the posterior distribution, because successive models from MCMC sampling usually share great similarity, and may cause bias for resolution analysis. The dashed green line is the true velocity model, where the seismic velocity at Moho changes abruptly from 7 km s⁻¹ of the lower crust to 8 km s⁻¹ in the upper mantle. The posterior distribution of the 1-D velocity model has higher density around the true model for both of the two parametrizations (Figs 6a and c); the Moho is clearly defined with the correct depth by most of the models in the posterior ensembles. As expected, the Moho structure is better defined using the hybrid model parametrization (compare Fig. 6c with Fig. 6a), since the step function for velocities from 9 km to greater depth better describes the abrupt velocity change at Moho.

Figs 6(b) and (d) show the posterior distribution for the depths of velocity nodes from the posterior models. The velocity model between two neighbouring nodes can be interpreted as a layer, and each velocity node can be viewed as a velocity interface or discontinuity. A great portion of velocity nodes is located where there are velocity discontinuities, for example, between 9 and 10 km depth. There is a

spike around 9.6 km depth in Fig. 6(d), which is consistent with the Moho interface in Fig. 6(c). Interestingly, the interface probability has another spike in the upper mantle (around 10.5 km), which corresponds to a velocity transition (Fig. 6c) constrained by refracted Pn wave.

The posterior distribution shows the non-uniqueness for seismic inversion problem. Figs 7(a) and (c) show the derived mean and the maximum posterior probability (MAP) models for the two parametrizations, respectively. Both the simulated (from the MAP model) seismograms (Figs 7b and d) from the two model parametrizations are visually identical with those from the true model for the offset of interest, which indicates the non-uniqueness of FWI problem. For a better understanding of PmP responses for different models, we also compute the seismograms for small offsets (Fig. 8), which are not included for inversion. The PmP amplitudes are much weaker than the true data in Fig. 8(b) from the Moho model with a transition zone, which indicates that the non-uniqueness can be reduced when there are more scattering angles in the data for illumination.

Figs 9(a) and (b) show a comparison of the prior and posterior distributions for the number of velocity nodes (model ‘dimension’ in the context of trans-dimension MCMC) using the two parametriza-

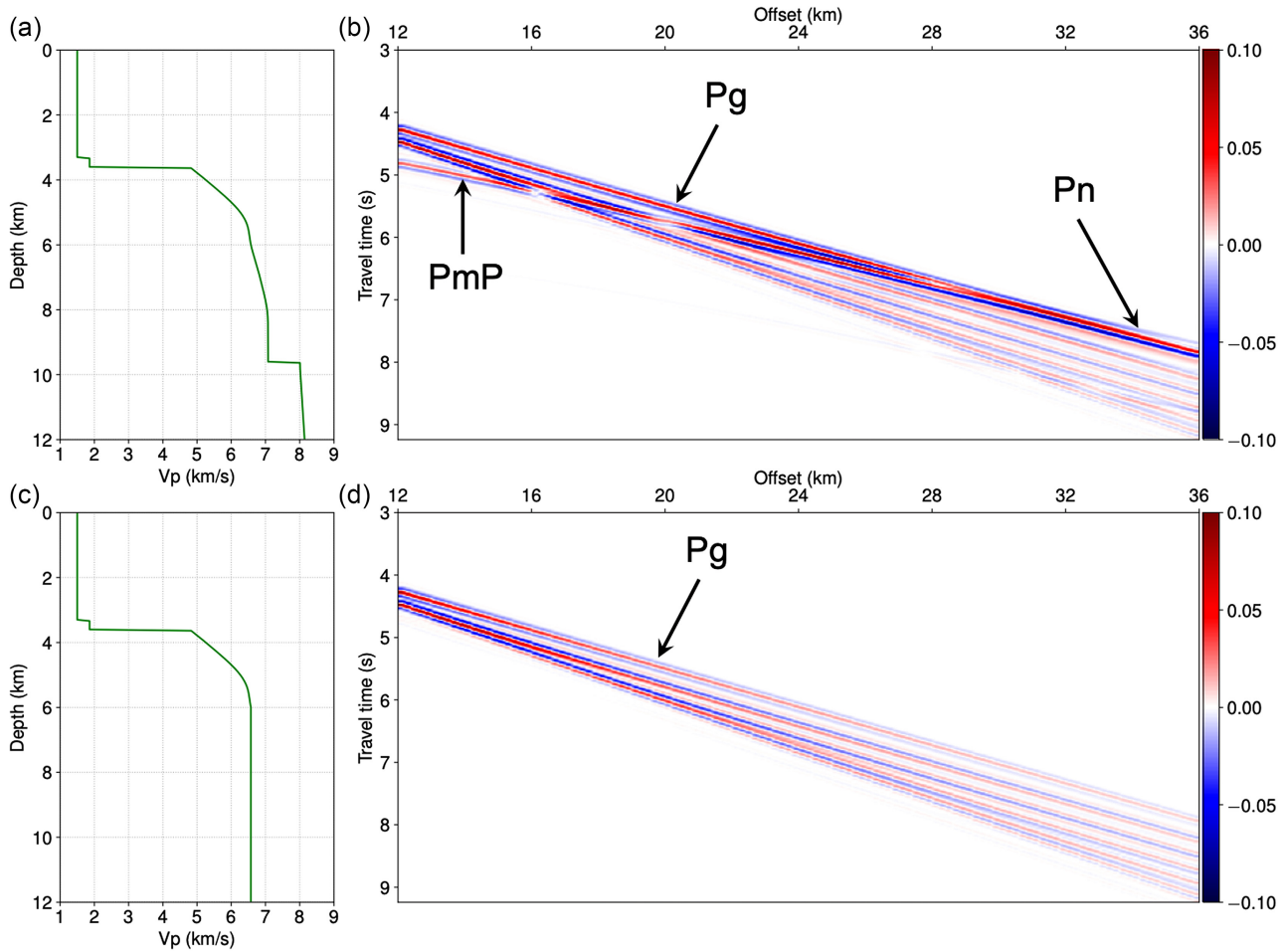


Figure 4. (b) A common OBS gathers using the velocity model of Fig. 2(a) (also plotted in (a)); and (d) a common OBS gathers using the velocity model in (c), which is the model at the first iteration of the trans-dimensional MCMC FWI. Three seismic phases are identified, including ‘ P_g ’ (turning wave), ‘ P_mP ’ (Moho reflection) and ‘ P_n ’ (refraction wave from Moho). The seismic data is simulated with a finite-difference solution of the 2-D acoustic wave equation (eq. 4).

tions. The number of velocity nodes in the posterior distribution is centred around 8 for the cubic spline interpolation; fewer velocity nodes are preferred in this example for defining the posterior velocity models using the hybrid model parametrization, with values around 5, because the sharp boundary in the Moho transition is better defined by using step function than cubic spline interpolation.

Fig. 10 shows the norm-2 data residuals as a function of iterations from 8 MCMC chains (randomly selected from the total 20 chains) for the cubic spline interpolation (Fig. 10a) and the hybrid model parametrization (Fig. 10c). The inversion converges after 60 000 iterations; Figs 10(b) and (d) show the zoomed-in data residual changes after convergence. The inferred velocity model ensembles in both Figs 6(a) and (c) generate predicted data with good match to the true data, with the model ambiguities (non-uniqueness) from the resolution limit of band-limited seismic data and the limited scattering angle range around critical offsets (P_mP phase in the field data usually appears only near critical angles). The time-series of velocity values at 10 km depth is shown in Fig. 11. As expected, there are large velocity fluctuations at early stages of the inversion, and the samples converge to the correct value (8.05 km s^{-1}) with increasing iterations. Both the data misfit and model parameters reach the state of equilibrium after 60 000 iterations, and a well-resolved structure appears in most of the sampled models.

3.2 Example 2

The second example tests the performance of trans-dimensional FWI for a Moho structure with a 0.8 km transition zone (Fig. 2b).

The posterior distribution density of the inferred velocity models (Figs 12a and c) indicates that trans-dimensional MCMC is able to find the global optima of velocity models for the strong nonlinear problem, starting from a model with constant velocity of 6.6 km s^{-1} (Fig. 4c) very far away from the true model; high probabilities (hot colour) follow the pattern of the true velocity model. The 0.8 km Moho transition zone with a smooth velocity gradient is defined well in Figs 12(a) and (c), using both parametrizations. The posterior for the depths of velocity nodes (Figs 12b and d) shows a large portion of nodes distributed close to and within the Moho transition zone, as expected, because large velocity changes require more nodes to define.

Fig. 13 shows the OBS gather from the true model (Fig. 13a) and the difference (Fig. 13b) with the data using the starting model (Fig. 4c) of trans-dimensional MCMC. There is waveform difference in turning waves because of velocity deviations in the lower crust and the Moho reflection. Figs 14(a) and (c) contain the velocity mean and the maximum a posterior probability (MAP) from the velocity model ensembles for the two model parametrizations,

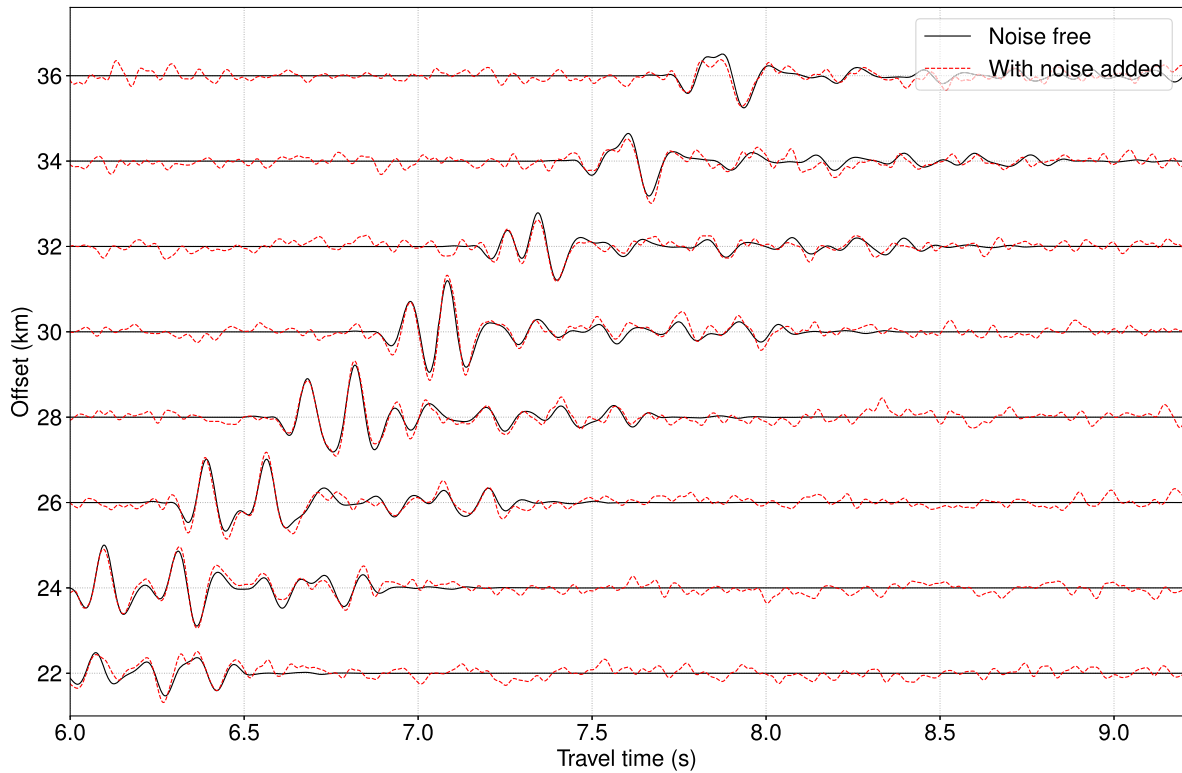


Figure 5. A group of selected seismic traces (solid black line) from Fig. 4(b); the seismic data after adding noise (dashed red line) is used as the ‘observed’ data for FWI.

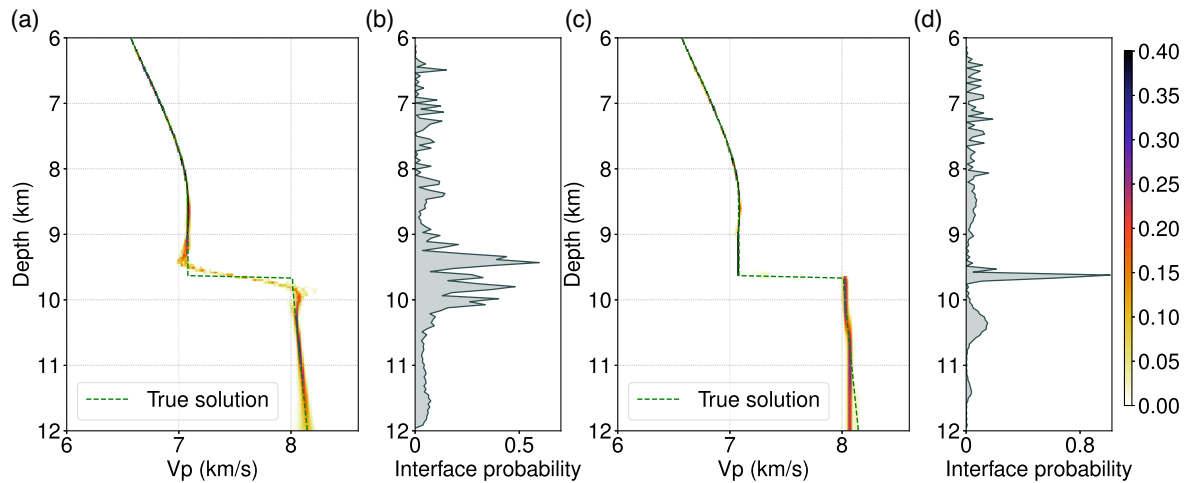


Figure 6. The posterior distribution density of the inferred velocity model from trans-dimensional FWI, using (a) the cubic spline interpolation and (c) the hybrid model parametrization, respectively; (b) and (d) shows the posterior distribution of the velocity node depths for these two parametrizations, respectively. The concentration of velocity nodes coincide with the distribution of velocity discontinuities in the true model. Prior for the velocity values are bounded between 6 and 8.6 km s⁻¹ with a mean of 7 km s⁻¹ and a standard deviation of 1 km s⁻¹ for a Gamma distribution.

respectively. Both the mean and MAP in Fig. 14(a) are almost identical to the true model. The MAP of the hybrid model parametrization (Fig. 14c) contains multiple layers in the Moho transition zone for describing the velocity changes in the true model; there is a great similarity between the mean and the true model in the transition zone. The difference (Figs 14b and d) between the predicted data using MAP and that from the true model is negligible; the difference in MAP models indicate the non-uniqueness of the FWI problem.

3.3 Example 3

The Moho model used here is shown in Fig. 2(c), with a transition zone from lower crust to upper mantle characterized by increasing velocity steps. Figs 15(a) and (c) show the posterior distribution of the model ensembles for the two model parametrizations, respectively. High probabilities locate around the true velocity model (the dashed green line), especially for the hybrid model parametrization. Instead of an exact match, the posterior distribution in Fig. 15(a) indicates that the inferred velocity models follow the velocity average

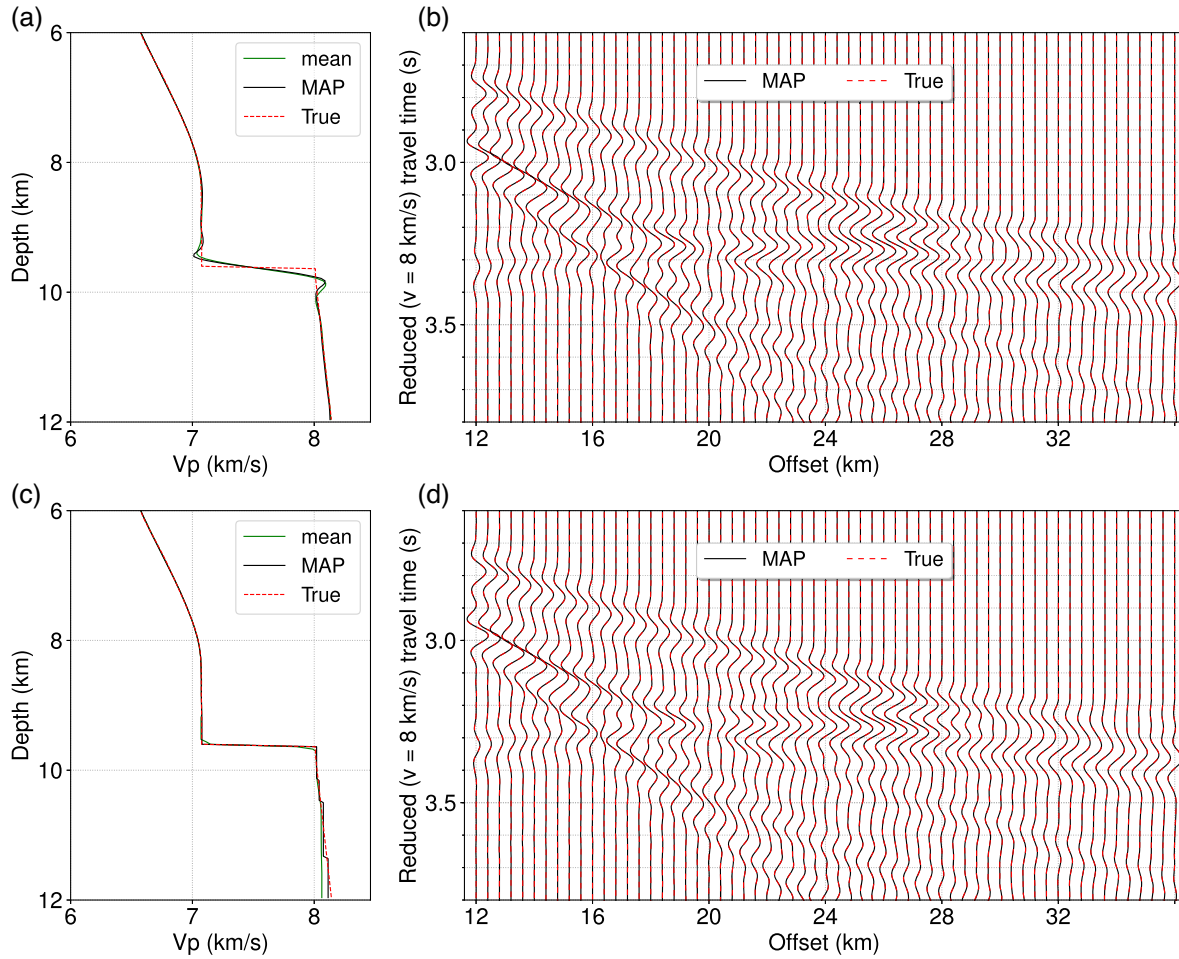


Figure 7. The mean and the maximum a posterior probability (MAP) models for (a) the cubic spline interpolation and (c) the hybrid model parametrization; the true model is plotted using the dashed red line. (b) and (d) show the corresponding OBS gathers from the true (dashed red line) model, and from the MAP (solid black line) models for the two parametrizations, respectively; the waveforms from the MAP and the true models are visually identical.

of the Moho transition zone, because it is difficult for cubic spline to present a series of increasing velocity steps. Similar with previous examples, the velocity nodes from the posterior models concentrate where there are velocity discontinuities, especially near and in the Moho transition zone (Figs 15b and d); moreover, we observe three peaks between 9.5 and 10.6 km for the posterior model ensemble using the hybrid model parametrization, with depths consistent with the increasing velocity steps in the transition zone of the true model.

Figs 16(a) and (c) show the mean and MAP of velocity models from the posterior ensemble, using the cubic spline interpolation and the hybrid model parametrization, respectively. Both the mean and MAP contain a smoothed version of the Moho transition in Fig. 16(a); a similar pattern of velocity steps in the MAP of Fig. 16(c) with the true model can be observed. Figs 16(b) and (d) show the corresponding differences between the simulated waveforms using the two MAP models and the true data, respectively. The small differences indicate a good waveform match for both of the two MAPs, which again, illustrates the non-uniqueness for seismic inversion problem arising from the limited frequency range (3–12 Hz for the synthetic study) and limited offsets for PmP arrivals.

3.4 Example 4

The final synthetic example uses seismic data from a Moho structure with low-velocity anomalies in the upper mantle (Fig. 2d). The low-velocity layers may be composed of intrusive gabbro. The posterior distribution (Figs 17a and c) of the inferred velocity models from wide-aperture seismic data are consistent with the true model (dashed green line), with high probability densities for both the Moho discontinuity and velocity anomalies in the upper mantle.

From the posterior distribution for the depths of velocity nodes, velocity nodes are more likely to be located where there are large velocity variations (Figs 17b and d); for the hybrid model parametrization where the step function is used from 9 km depth to upper mantle, there are five spikes (Fig. 17d) at depths which are consistent with the seismic velocity discontinuities in the true model.

One of the attractive properties for FWI using trans-dimensional MCMC is that, besides the physical properties of the subsurface, it is also able to infer the model complexity for defining the velocity–depth structure, therefore removing the preparation work for determining the number of unknowns before inversion. Fig. 18 provides the posterior distribution for the number of velocity nodes. Compared with those in Figs 9(a) and (b), the required number of velocity

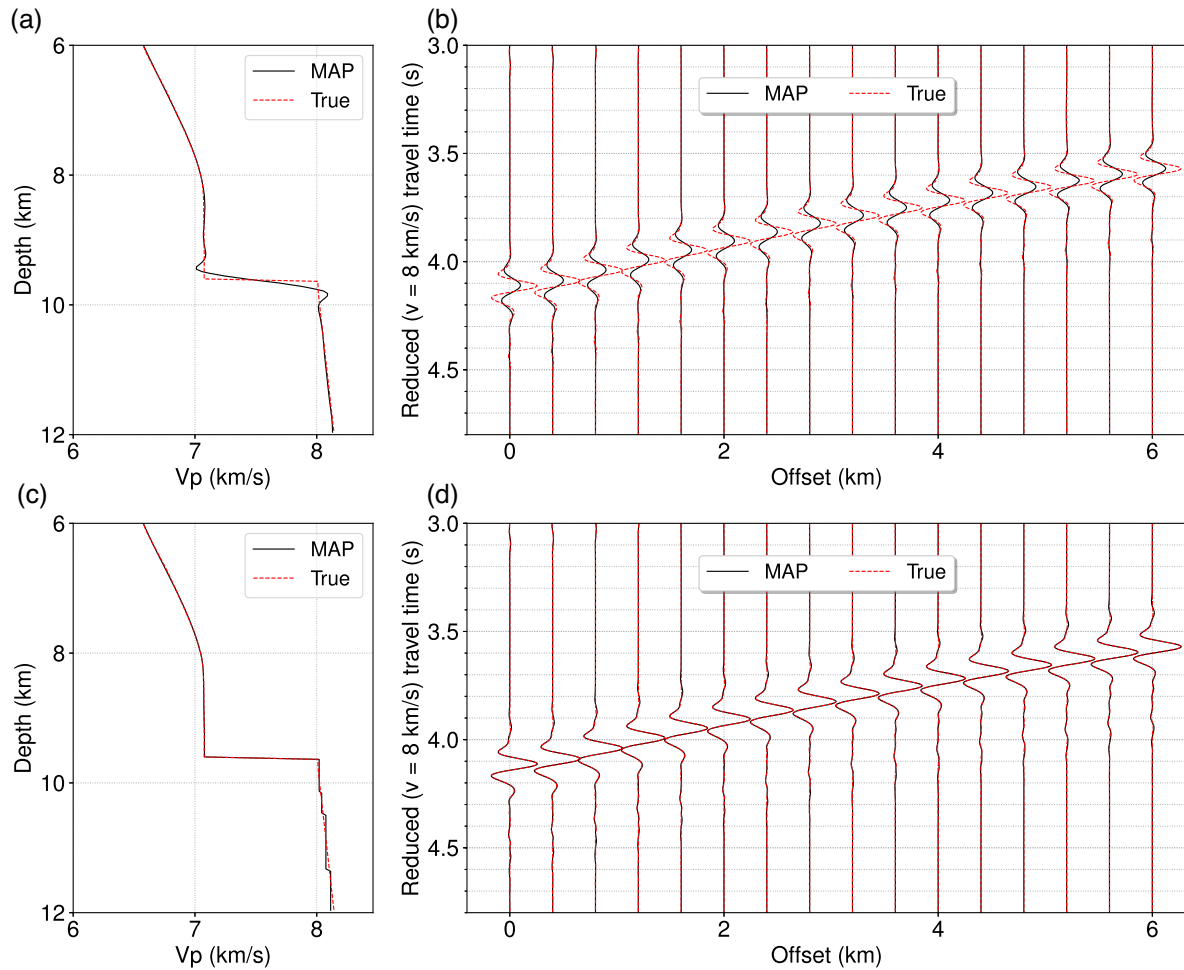


Figure 8. (a) and (c) plot the maximum a posteriori probability (MAP, solid black line) models from inversion using (a) the cubic spline interpolation and (c) the hybrid model parametrization, and the true model (dashed red line). (b) and (d) show the corresponding simulated OBS gathers from the MAP (solid black line) for the two parametrizations, respectively, and for the true (dashed red line) model. The amplitudes of MAP are weaker than the true data in (b), and those in (d) are visually identical for both the two parametrizations.

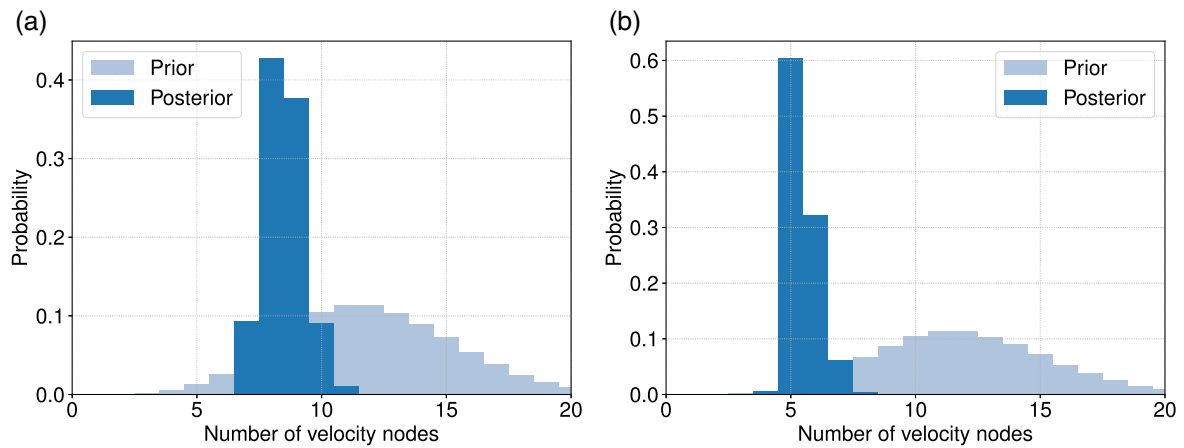


Figure 9. The prior and posterior probability for the number of velocity nodes using (a) the cubic spline interpolation and (b) the hybrid model parametrization.

nodes in the posterior models increases, because there are more velocity discontinuities in the velocity model of Fig. 2(d) than that in Fig. 2(a).

Fig. 19 shows three randomly selected velocity models from the posterior ensemble, and the comparison between the correspond-

ing predicted (solid black line) and the observed (dashed red line) seismograms, for the cubic spline interpolation (Figs 19a–c) and the hybrid model parametrization (Figs 19d–f), respectively. Velocity uncertainty and the non-uniqueness of inversion problem can be observed, as different velocity models predict similar synthetic

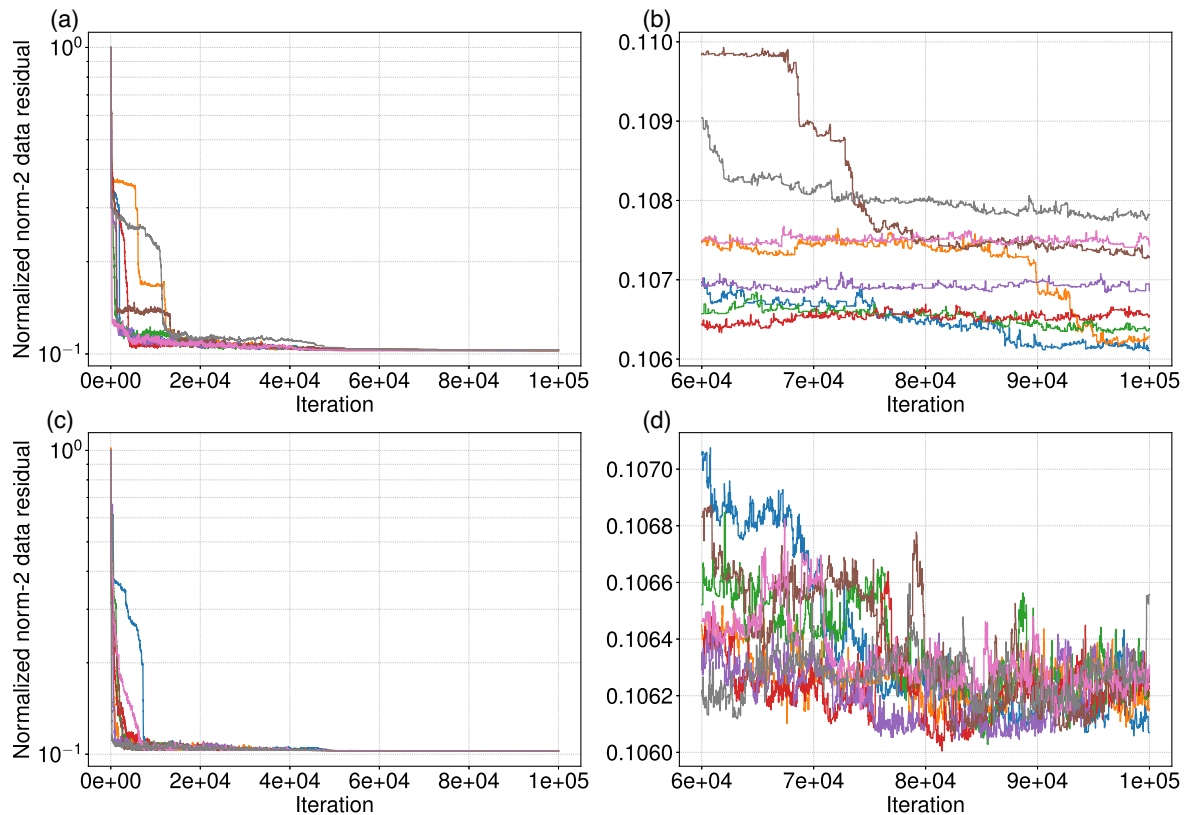


Figure 10. The norm-2 data residual as a function of iterations for (a and b) the cubic spline interpolation and (c and d) the hybrid model parametrization. The inversion converges after 60 000 iterations. Different colour indicates different independent MCMC chains; 8 chains (randomly selected out of 20 chains) are plotted. (b) and (d) show the data residuals in the post-burn-in stage for the two parametrizations, respectively.

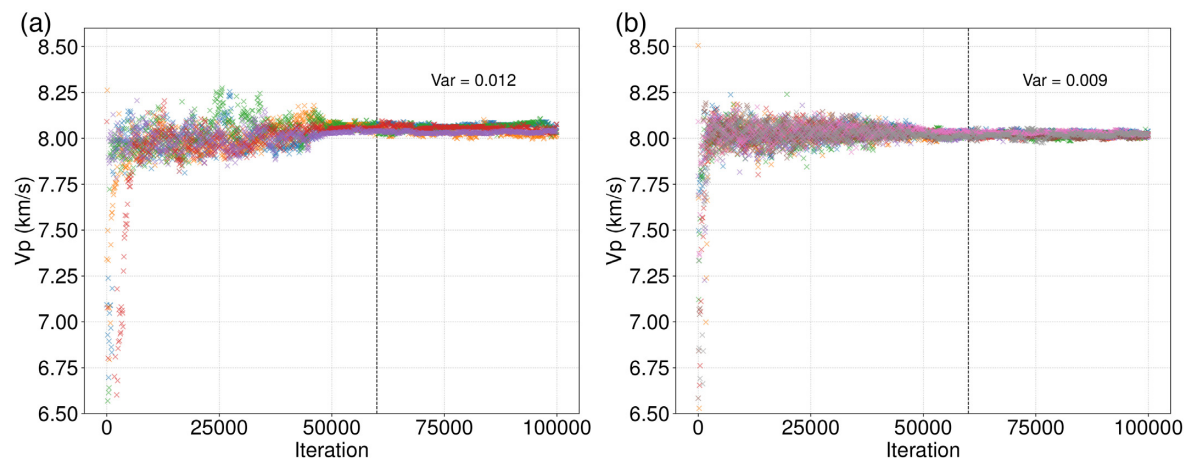


Figure 11. The evolution of velocity value at depth 10 km as a function of iterations for (a) the cubic spline interpolation and (b) the hybrid model parametrization. The vertical dashed black line is located at 60 000 iterations, where MCMC chains are considered reaching convergence. The variance of the post-burn-in stage for the two parametrizations is also labelled in the plots.

seismograms matching those from the true model very well. There are trade-offs between velocity and interface depths (layer width); for example, a lower estimated velocity may be compensated by a relatively thicker layer (Fig. 19d) because of the frequency range and acquisition we used.

4 TRANS-DIMENSIONAL FWI FOR MOHO STRUCTURE USING OBS DATA FROM THE MID-ATLANTIC OCEAN

Synthetic studies using four different Moho scenarios demonstrate that, with little prior information, trans-dimensional MCMC is able

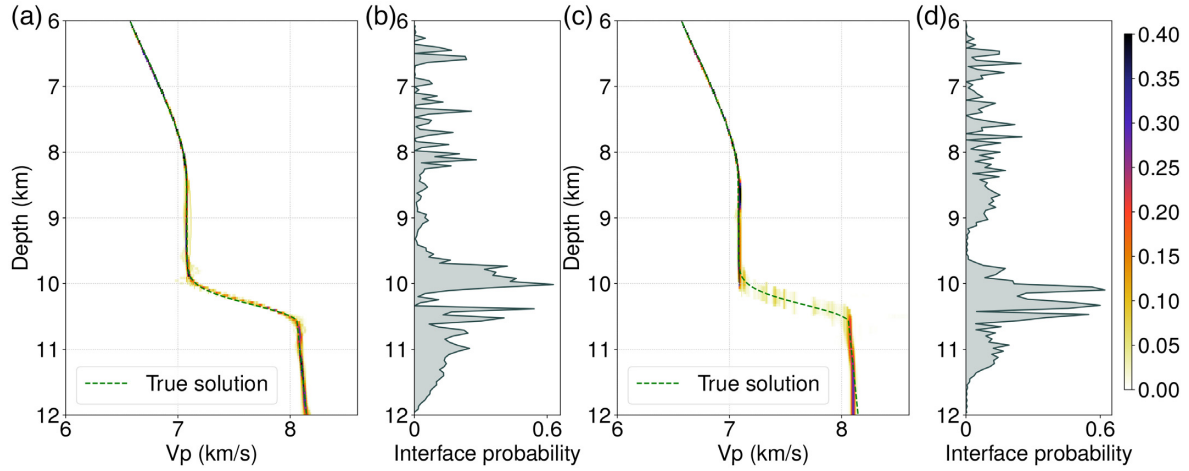


Figure 12. The posterior distribution density of the inferred velocity model from trans-dimensional FWI, using (a) the cubic spline interpolation and (c) the hybrid model parametrization, respectively; (b) and (d) shows the posterior distribution of the velocity node depths for these two parametrizations, respectively. The concentration of velocity nodes coincide with the distribution of velocity discontinuities in the true model.

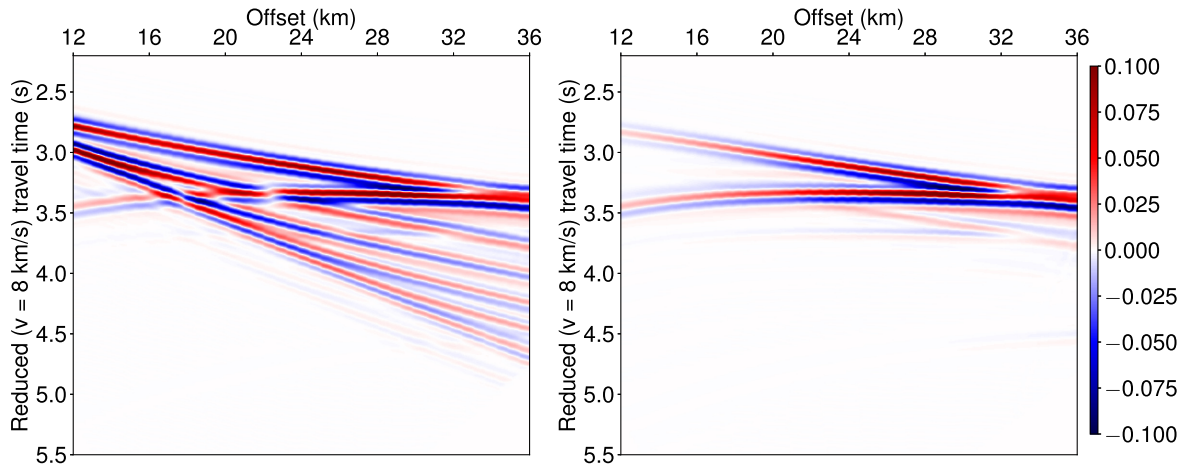


Figure 13. (a) The seismic data simulated using the true model in Fig. 2(b); (b) the difference with the data using the velocity model (Fig. 4c) at the first iteration of trans-dimensional MCMC. The colour scale is the same with Fig. 4. Note that a reduced velocity of 8 km s^{-1} has been applied to the data.

to provide a probabilistic solution for a strong nonlinear FWI problem using wide-aperture seismic data, including the posterior distribution for the model complexity (the number of layers) and physical parameters. In this section we invert for the Moho structure using a field OBS data from the Mid-Atlantic Ocean.

4.1 OBS data from the Mid-Atlantic Ocean and pre-processing

The study region locates in the equatorial Atlantic Ocean over the crust formed at the MAR (Fig. 20). The field OBS data were acquired during the LITHOS-iLAB cruise in 2017 November–December on board the R/V Maria S. Merian to study the age-dependent features of the upper lithosphere from 0–50 Ma in the study region. There are 71 instruments in total consisting of 55 OBSs and 16 OBH (ocean bottom hydrophone), which were deployed along a 1100 km long transect with a variable spacing of 10–20 km, for recording long-offset wide-angle seismic refractions and reflections. The seismic profile (around 1000 km) mainly lies on the African plate (0–50 Ma) and crosses the ridge-axis on to the

South American plate (0–2 Ma) for less than 75 km. All OBS were equipped with a hydrophone (measuring pressure) and three geophones (measuring vertical and horizontal displacements) whereas the OBH measured only the pressure. The data was sampled at 250 Hz for all the instruments. The active seismic source used in the survey comprised of 6 G-gun clusters (12 guns) configured as two subarrays with a total volume of 5440 cu in, which was towed at 7.5 m in depth and fired around every 400 m along the profile giving a total of 2735 shots. The relatively larger shot spacing was chosen to reduce the noise from neighbouring shots and hence enhancing the SNR for farther offsets.

Since the OBS is deployed from the sea surface, the actual OBS location is likely to be drifted away from the deployment position at surface, because of, for example, ocean currents. Therefore before processing the data, all OBS were repositioned using the direct water wave arrivals up to 5 km offset on both sides of the OBS.

We limit the data pre-processing to a minimum to keep the waveform information, including a bandpass filtering of 4–10 Hz, and 3-D to 2-D transformation [multiplying the amplitudes of field data by \sqrt{t} (where t is the two-way traveltime) and convolving with $1/\sqrt{t}$; Pica *et al.* 1990; Forbriger *et al.* 2014] because 2-D wave

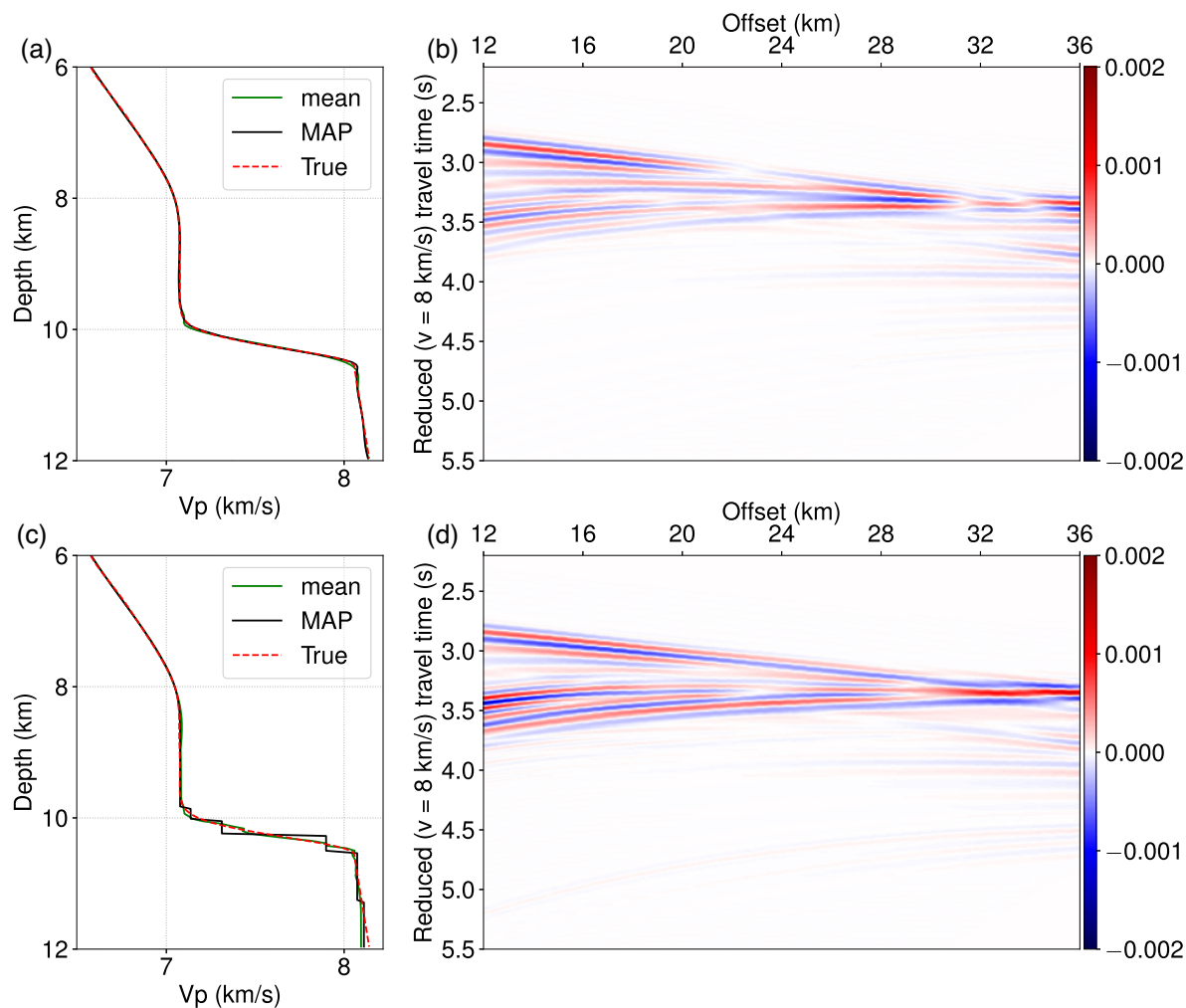


Figure 14. The mean and the maximum a posteriori probability (MAP) models for (a) the cubic spline interpolation and (c) the hybrid model parametrization; the true model is plotted using the dashed red line. (b) and (d) show the corresponding data difference between the true data and data simulated using MAP for the two model parametrizations; the difference is negligible. Note that a reduced velocity of 8 km s^{-1} has been applied to the data, and the colour scale here is smaller than that in Fig. 4 for visualizing the difference.

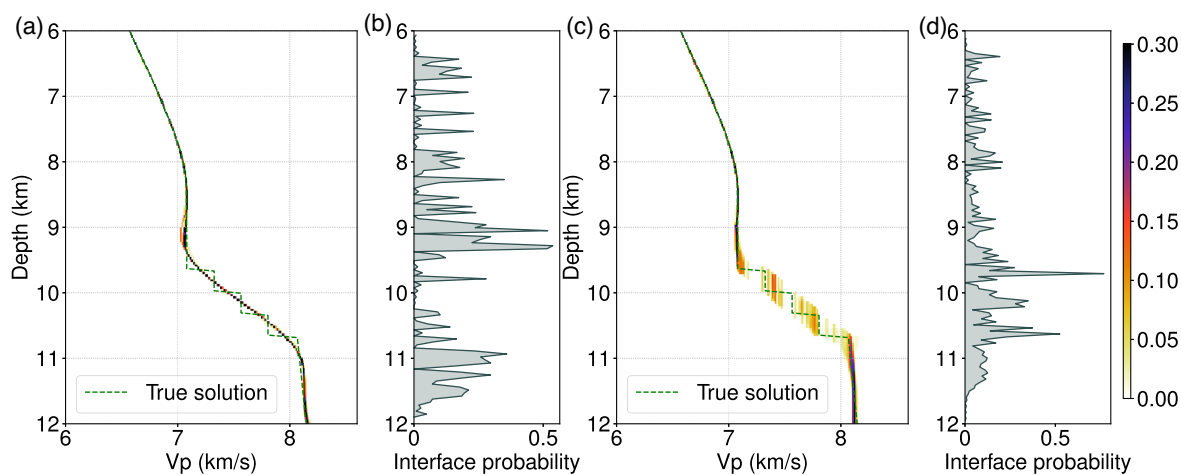


Figure 15. The posterior distribution density of the inferred velocity model from trans-dimensional FWI, using (a) the cubic spline interpolation and (c) the hybrid model parametrization, respectively; (b) and (d) shows the posterior distribution of the velocity node depths for these two parametrizations, respectively. The concentration of velocity nodes coincide with the distribution of velocity discontinuities in the true model.

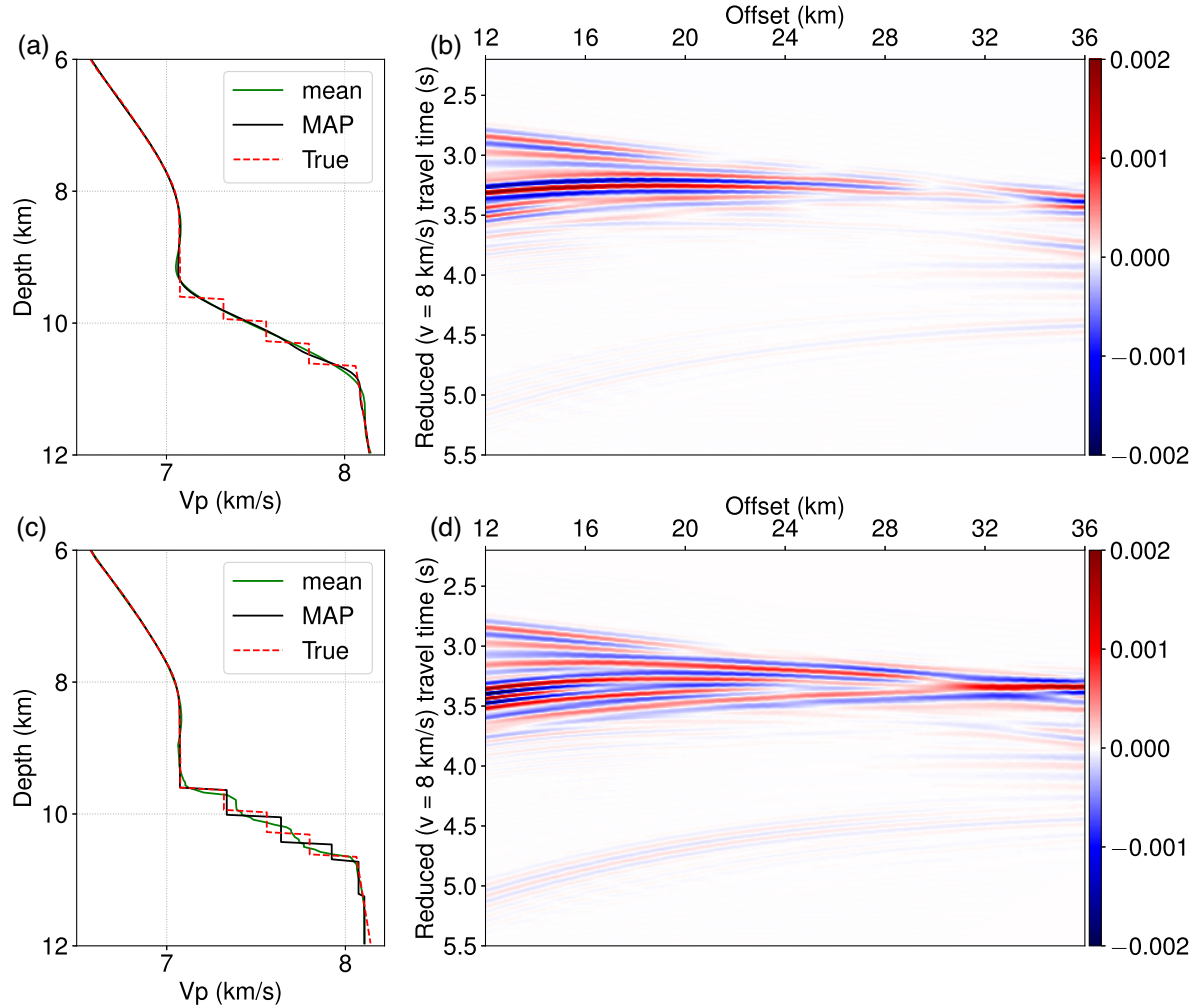


Figure 16. The mean and the maximum a posteriori probability (MAP) models for (a) the cubic spline interpolation and (c) the hybrid model parametrization; the true model is plotted using the dashed red line. (b) and (d) show the corresponding data difference between the true data and data simulated using MAP for the two model parametrizations. Note that a reduced velocity of 8 km s^{-1} has been applied to the data, and the colour scale here is smaller than that in Fig. 4 for visualizing the difference.

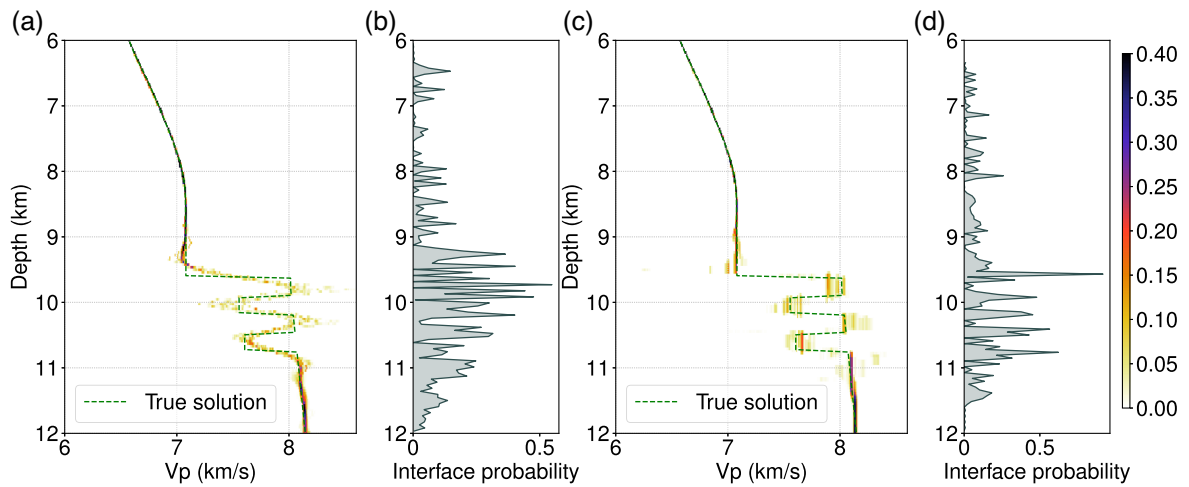


Figure 17. The posterior distribution density of the inferred velocity model from trans-dimensional FWI, using (a) the cubic spline interpolation and (c) the hybrid model parametrization, respectively; (b) and (d) shows the posterior distribution of the velocity node depths for these two parametrizations, respectively. The concentration of velocity nodes coincide with the distribution of velocity discontinuities in the true model.

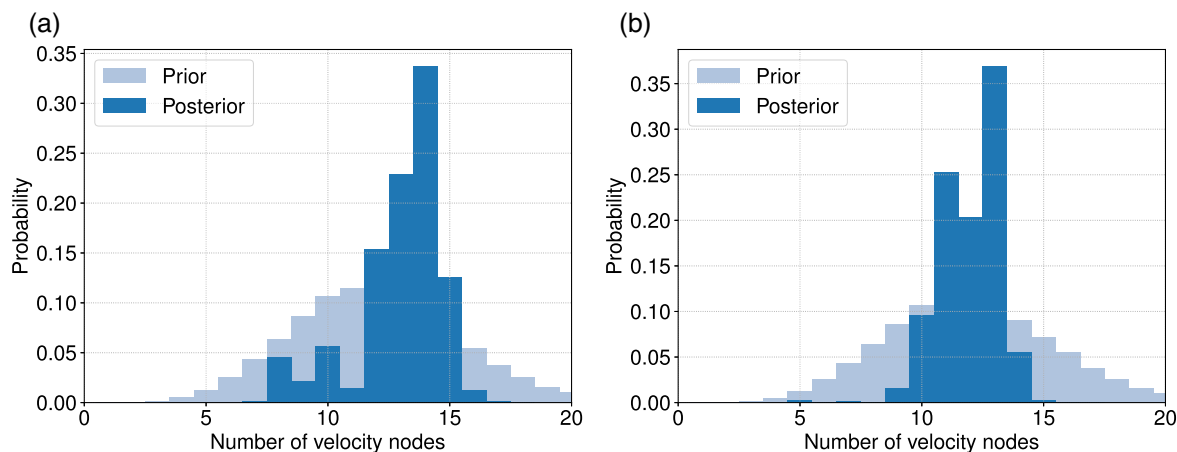


Figure 18. The prior and posterior probability for the number of velocity nodes using (a) the cubic spline interpolation and (b) the hybrid model parametrization. Compared with those in Fig. 9, the required number of velocity nodes increases because of increasing model complexity.

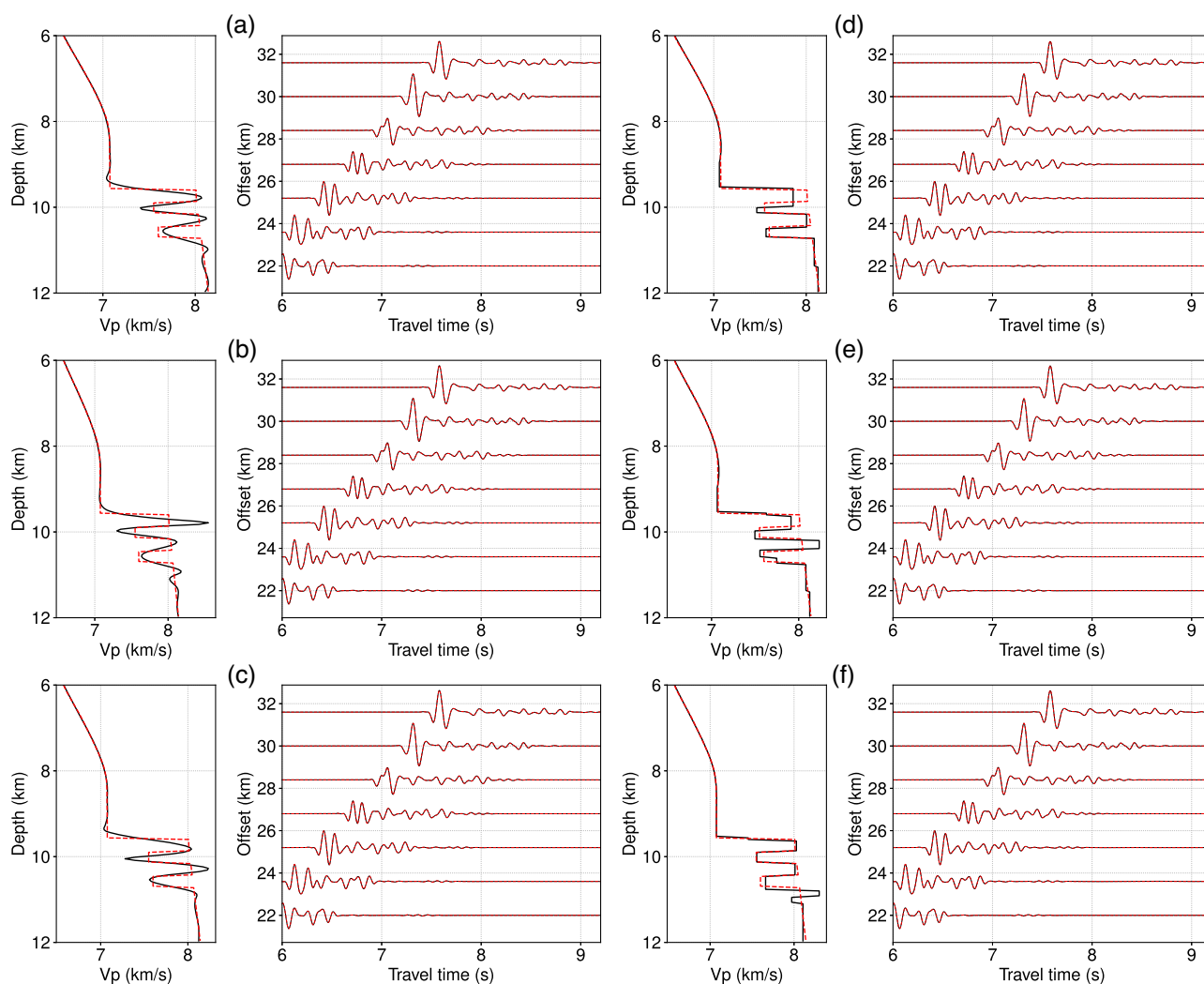


Figure 19. Randomly selected velocity models and the corresponding comparison between the predicted data and the observed data, for (a–c) the cubic spline interpolation and (d–f) the hybrid model parametrization. In the velocity model plots, the dashed red line is the true model, and the solid black line is the velocity models from trans-dimensional FWI; in the seismic trace plots, the dashed red lines are the seismic waveform from the true model, and the solid black lines are for the seismic data from inverted models.

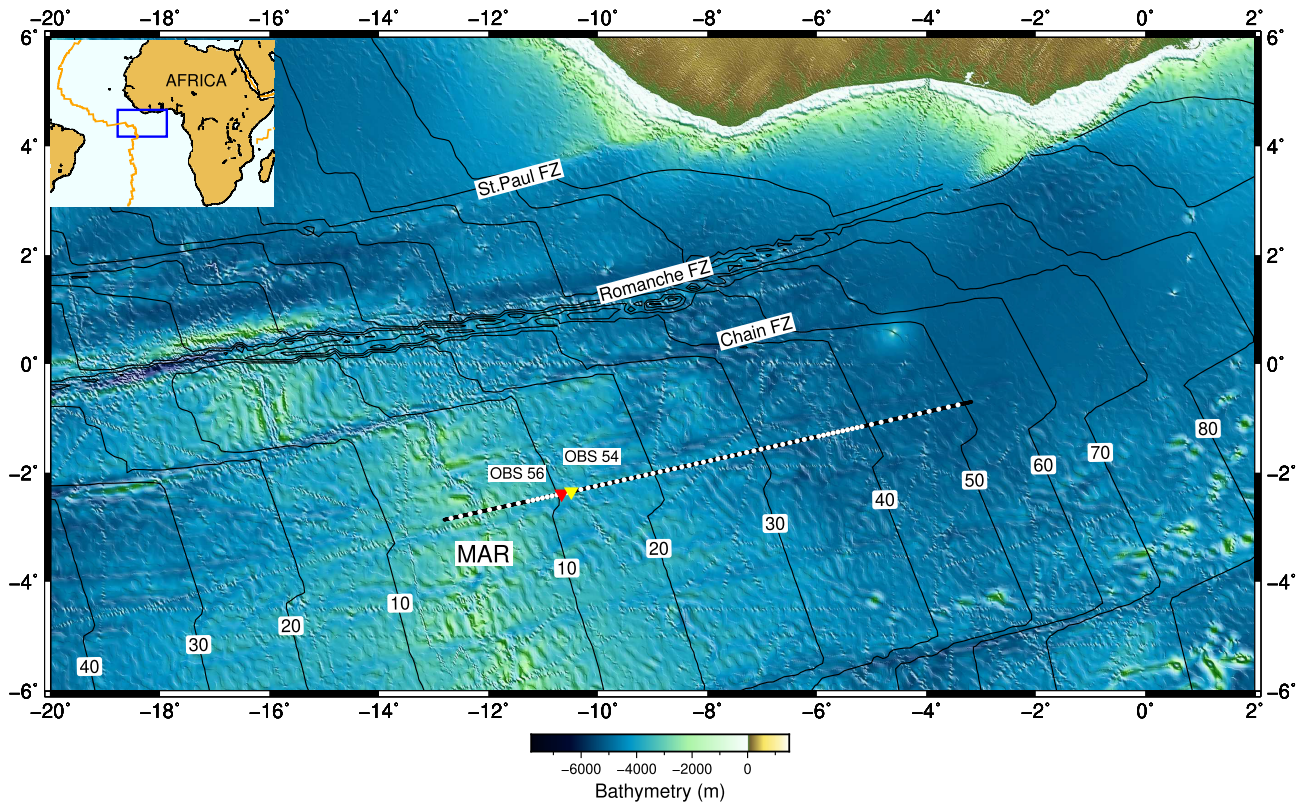


Figure 20. The OBS seismic survey over the crust formed at the Mid-Atlantic Ridge (MAR) in the Mid-Atlantic Ocean. The black dots refer to the locations of a total of 71 OBS instruments. The red triangle refers to the location for OBS 56, and the yellow one nearby refers to OBS 54.

equation modelling is used for seismic wave simulation at each iteration of inversion. A predictive deconvolution is applied to suppress the bubble effects from the air gun sources.

Traveltime tomography (Van Avendonk *et al.* 1998, 2004; Vaddineni *et al.* 2018) is first applied to the OBS gathers for estimating a P wave velocity model (Fig. 21). We consider that the velocity model up to 9 km depth is constrained from linearized inversion using, for example, the P_g phase (approximately a linear relation for velocity and waveforms of different offsets), and only invert for the velocity structure at depths between 9 and 12 km, where the Moho is likely to be present. The sparse 1-D model parametrization (no lateral heterogeneities) provide a good approximation for the velocity model within the study depth (Fig. 21, the model region between the two vertical blue lines), for OBS 56 with maximum offset 40 km, therefore we consider it a reasonable assumption for trans-dimensional FWI. There are relatively strong lateral heterogeneities in the velocity model from tomography, therefore we use 2-D acoustic wave equation modelling for a better modelling of wave physics in the crust than 1-D simulation.

We choose the PmP (denoted by dashed red line) and Pn (denoted by dashed black line) phases from the common OBS gather of OBS 56 (Fig. 22) as the observed data for estimating Moho discontinuities using trans-dimensional FWI. The offset range we used is between 24 to 36 km. OBS 56 is chosen mainly because of a good SNR at far offsets.

4.2 Source wavelet estimation

FWI requires a comparison of field and synthetic seismograms at each iteration, and the synthetic seismograms are obtained by full

waveform forward modelling using a numerical solution of the wave equation. The velocity model estimated by FWI is sensitive to the source wavelet, as the simulated seismograms are determined by both the source signature and subsurface model. Thus it is important to obtain a reliable estimate for the source wavelet.

The direct arrival from the source position (water waves) and the seafloor reflection overlaps with each other in the common OBS gather (Fig. 22), along with strong scatterings and reflections from sedimentary layers close to the seafloor, making it difficult to isolate a reliable source wavelet. One option is to use the P_g phase. There is no interaction between turning waves and reflectors, the waveform of the P_g phase can potentially keep the source signature. However, the nearest-offset P_g phase that can be picked from OBS 56 data is around 5 km; lateral heterogeneities in the crust can damage the source signature during propagation. Alternatively, another option is the near-offset free-surface multiples; we observe a good separation between the free-surface multiples of water waves and reflections, as indicated by the yellow box in Fig. 22. Therefore the near-offset free-surface multiples of water waves are used for estimating source wavelet.

The dashed lines in Fig. 23(a) plot the recorded near-offset free-surface multiples of water waves, with offset ranges from -0.49 to 1.14 km. The water waves are filtered into the frequency range of 4–40 Hz. After trace alignment (the solid black lines of Fig. 23a), we observe great similarities between traces; the aligned trace are stacked together (the solid red line), which is then bandpass filtered (4–10 Hz, the green line). Because the reflection coefficient is -1 at the free surface, the waveform in green (Fig. 23a) is multiplied by -1 before being used as the source wavelet.

We use the estimated source wavelet for forward modelling using acoustic wave equation, with the velocity model in Fig. 21.

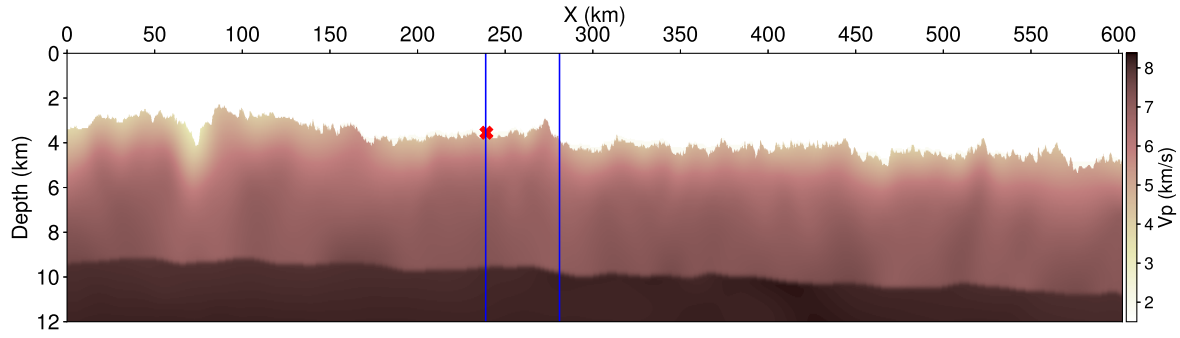


Figure 21. The velocity model from traveltime tomography. The red cross refers to the location of OBS 56. The model between the two vertical blue lines is used for trans-dimensional FWI using seismic data of OBS 56; the velocity model to 9 km depth is kept unchanged during FWI, and we only invert for the velocity structure from 9 to 12 km in depth using 1-D model assumption.

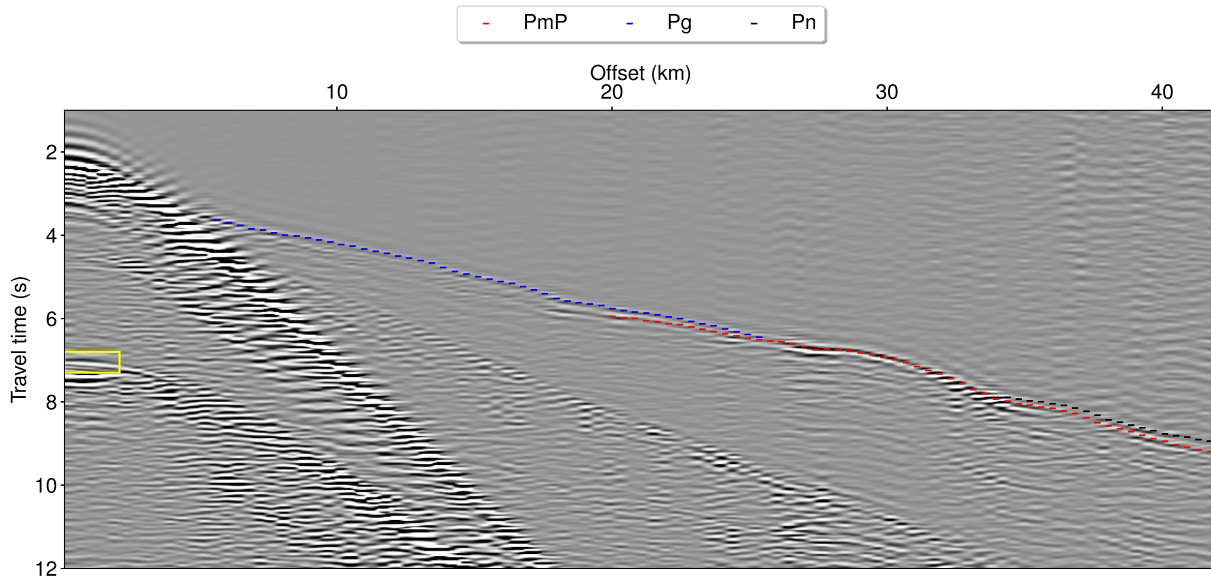


Figure 22. Seismic data for OBS 56, where the ‘PmP’, ‘Pg’ and ‘Pn’ phases are indicated by the dashed red, blue and black lines. The yellow box at near offsets around traveltime of 7 s shows a good separation between free-surface multiples of the water waves and subsurface reflections.

Fig. 23(b) shows a good match of the Pg phase between the field and the simulated data, which indicates that the estimated source wavelet provides a reliable source signature for FWI. We only need to adjust the amplitude of the synthetic data by multiplying a scaling factor, which can be obtained by comparing the amplitude of seismic waveforms between the synthetic and field data.

4.3 Prior for the field data

For inversion using seismic data from OBS 56, weak prior information is used. A Gamma distribution with a mean velocity value of 7.8 km s^{-1} , and a standard deviation of 0.8 km s^{-1} are used as prior distribution for the velocities. We set the maximum and minimum velocities at 6.6 and 9 km s^{-1} , respectively, to avoid unrealistic velocity values.

A symmetric Dirichlet prior distribution with concentration parameter of 1 is used for the vector of layer thickness, with the values summed equal to the maximum model depth, which we set to 3 km.

The prior over the number of velocity nodes is defined by a Poisson distribution. In this study, we use a value of 8 as the mean

for the prior distribution of the node numbers, and set the maximum number to be 20.

Density is also incorporated into seismic forward modelling to avoid the overestimation of velocity for the subsurface model. We constrain the density from velocity using the following relationship (Shipp & Singh 2002):

$$\rho = \begin{cases} 1000 & \text{if } V_p \leq 1.5, \\ 2351 - 7497 V_p^{-4.656} & \text{if } 1.5 < V_p < 2.2, \\ 1740 V_p^{0.25} & \text{if } V_p \geq 2.2, \end{cases} \quad (9)$$

where V_p is the P -wave velocity with unit of km s^{-1} , and ρ is the density with unit of kg m^{-3} .

4.4 Inversion results

The seismic traces between the offset range of 24 and 36 km are used as data constraints, which contain mainly PmP reflections around the critical angle and some of the Pn phase. We use 20 independent trans-dimensional MCMC chains to explore the model space. The inferred parameters include the number of velocity nodes for describing the model complexity, and the physical parameters

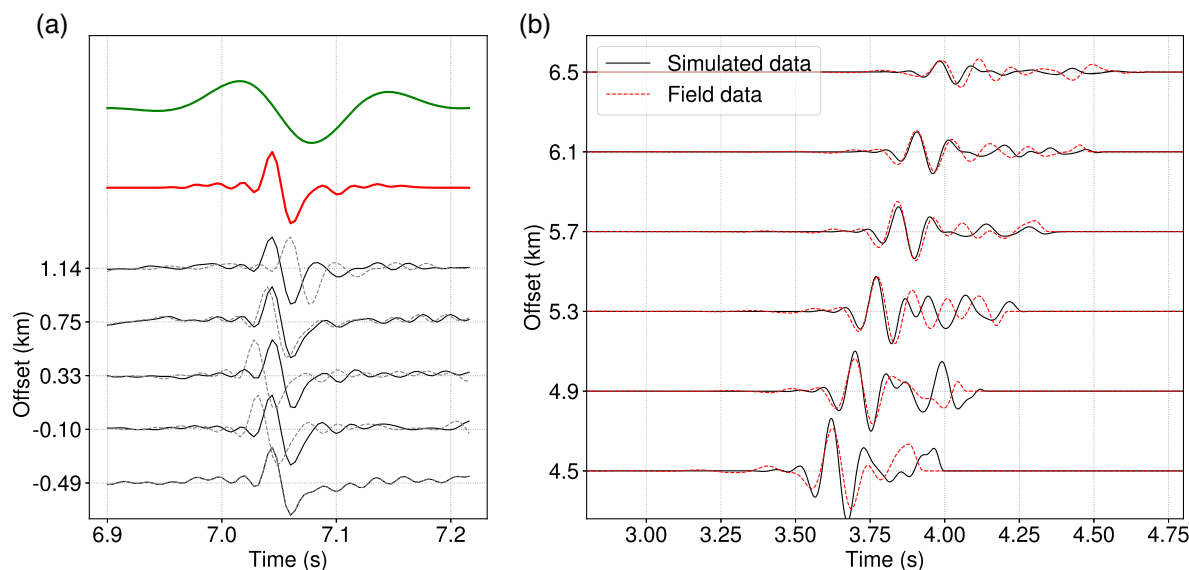


Figure 23. (a) shows the seismic traces (dashed line lines) of free-surface multiple water waves at near offsets from the area circled by the yellow box in Fig. 22; the red line shows the stacking result after trace alignment (solid black lines). The green line shows the bandpass filtered (4–10 Hz) waveform from the stacking result. Because the reflection coefficient is -1 at free surface, the waveform of green colour multiplied by -1 is used as the source wavelet. (b) shows the comparison between the simulated data using the estimated source wavelet and the field data for the Pg phase.

for defining the velocity discontinuities, including the velocity value and depth; the 1-D velocity models can then be interpolated using the sparsely distributed velocity nodes. There are 120 000 iterations for each MCMC chain. When analysing results, the first 80 000 iterations are considered as the ‘burn-in’ stage, and the sampled velocity models are discarded. Velocity models from the last 40 000 iterations of each chain are used as the posterior velocity model ensemble.

Fig. 24 shows the posterior distributions of the velocity models and the distribution of velocity node (discontinuities) depths for the cubic spline interpolation (Figs 24a and b) and the hybrid model parametrization (Figs 24c and d). Fig. 24(a) indicates a Moho transition zone of about 0.7 km width with increasing velocities, and Fig. 24(c) prefers a Moho interface located at about 9.4 km depth, with velocity changing from about 7 km s⁻¹ at lower crust to 8 km s⁻¹ at upper mantle. The velocity uncertainty increases with increasing depths because of the lack of constraints for greater depths.

It is common to use one seismic gather for FWI with a 1-D assumption, especially in the framework of Bayesian theorem (Ray *et al.* 2016). For the completeness of this study, we also invert for Moho using two OBS gathers, with the additional OBS gather indicated by the yellow triangle in Fig. 20. We use the positive offsets for OBS 56 and negative offsets for OBS 54, with a mid-point between the two about 10 km. While a better resolution is usually expected with more data for illuminating the target, the posterior distribution (Fig. 25) from trans-dimensional FWI using two OBS gathers shows great similarity with the inversion results (Fig. 24) using only OBS 56; there is no obvious resolution improvement. Possible reasons include that, using additional OBS gathers nearby may not necessarily enrich the effective scattering angles for illuminating the Moho, because for both the two OBS gathers, the PmP arrivals appear only at far offsets around the critical angle. Secondly, a 1-D velocity–depth model for describing the Moho structure may be a coarse approximation for complex geological settings; using additional OBS gathers increase the data coverage, therefore may increase the discrepancies between the 1-D assumption and the

3-D subsurface. Using more OBS gathers for inversion makes it necessary to consider lateral heterogeneities by employing a more sophisticated model parametrization.

Fig. 26 shows the MAP models from the posterior velocity ensembles for both of the two model parametrizations, and their corresponding comparison with the field OBS data. The predicted seismic waveforms (solid black lines in Figs 26b and d) from the two MAP models are very similar and both match the observed data (dashed red lines) well. Whereas there is ambiguities in determining the preferred velocity model, because of the limited resolution of the band-limited (4–10 Hz for the field data) seismic data and the limited available offsets that is used for this study, there are substantial velocity changes between 9.3 and 9.6 km, supported by the posterior distribution of velocity models (Figs 24a and c) and the distribution for depths of velocity nodes (Figs 24b and d).

As there are two inverted models (Fig. 24) for the field data, it may be reasonable to follow Occam’s razor rule for choosing a simpler model; on the other hand, we notice that Moho reflection can only be observed in the far offsets (larger than 20 km). The missing of Moho reflection in the near to intermediate offsets may indicate a weak reflection for smaller scattering angles. Considering that PmP from a smooth Moho transition can be identical at far offsets with that from Moho with sharp boundary, but the amplitude can be much weaker at small offsets (see Figs 7 and 8), therefore it may be reasonable to prefer the Moho model with a 0.7 km transition zone. However, it is possible that PmP was recorded but interfering with other waves in the near offsets. It can be helpful to acquire new wide-aperture seismic data with higher S/N and wide frequency spectrum for improving the resolution of Moho.

The results of this study indicate the possibility of a layered structure in the vicinity of Moho (Fig. 27) at slow spreading MAR. Although the two MAP models are different, they suggest a 1.5–2 km thick transition zone in the upper mantle. One of the models shows lower and higher velocity layers in the upper mantle, where low-velocity layers in the mantle can be composed of Gabbro layers, and the high-velocity layers may indicate Peridotite; the other model

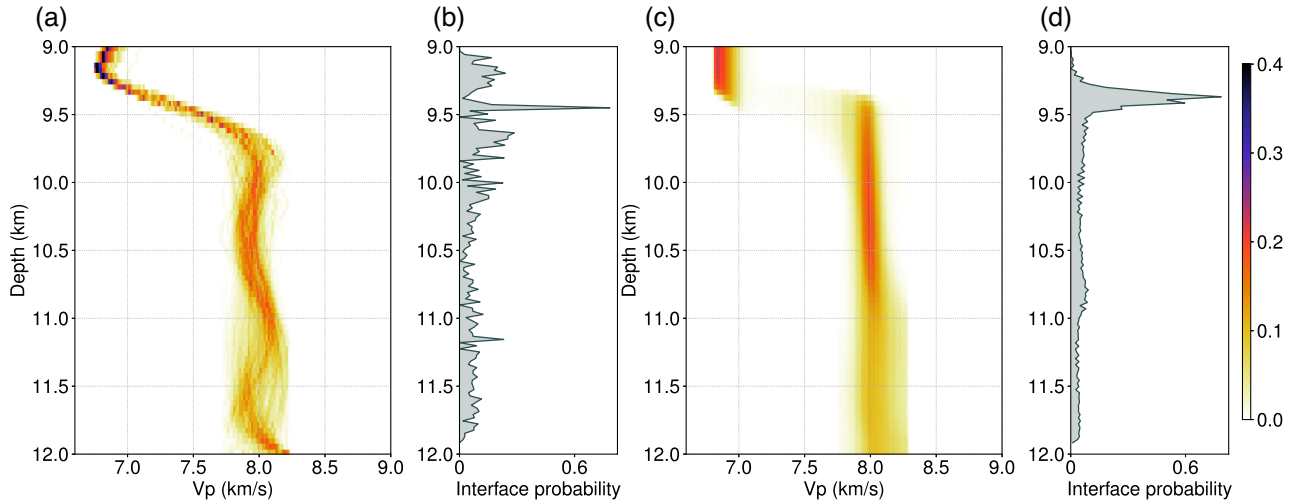


Figure 24. The posterior distribution density of the inferred velocity model from trans-dimensional FWI using one common OBS gather (OBS 56), using (a) the cubic spline interpolation and (c) the hybrid model parametrization, respectively; (b) and (d) show the posterior distribution of the velocity node depths for these two parametrizations, respectively.

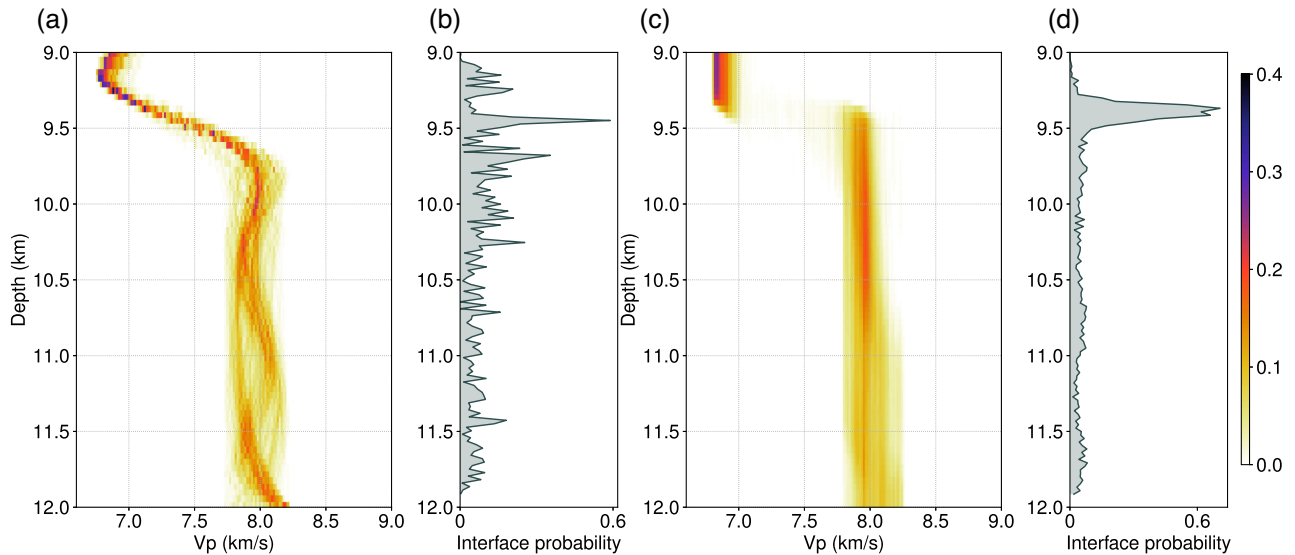


Figure 25. The posterior distribution density of the inferred velocity model from trans-dimensional FWI using two common OBS gathers (OBS 56 and 54), for (a) the cubic spline interpolation and (c) the hybrid model parametrization, respectively; (b) and (d) show the posterior distribution of the velocity node depths for these two parametrizations, respectively.

indicates increasing velocity layers in the upper mantle. These results are consistent with the absence of any strong seismic reflection image for slow spreading oceanic crust and observation from the Bay of Island Ophiolite (Brocher *et al.* 1985). The velocity structure in the relatively deeper part of the upper mantle is mainly constrained by the Pn phase, which has limited penetration depth. There are relatively large waveform misfits at very far offsets (start from 35.2 km), where the overlap of PmP and Pn phases appears (Fig. 22). The results can be improved using more sophisticated model parametrization for describing the possibly heterogeneous upper mantle, and more efficient model searching algorithms for practical application in future studies.

5 DISCUSSION

FWI of the PmP phase is rarely used for Moho model building, because of the strong nonlinearity observed in the vicinity of the critical offsets, the waveform interference from Pg and Pn phases, and the relatively low SNR for PmP phase compared with seismic reflections from the upper crust (Grad *et al.* 2009; Hrubcová *et al.* 2013; Jian *et al.* 2017; Beller *et al.* 2018). Trans-dimensional MCMC provides a fully nonlinear approach for inverse problems (Sambridge *et al.* 2006). Unlike the common linearized inversion method that the solution is a single final model which strongly depends on the starting model, the trans-dimensional MCMC method

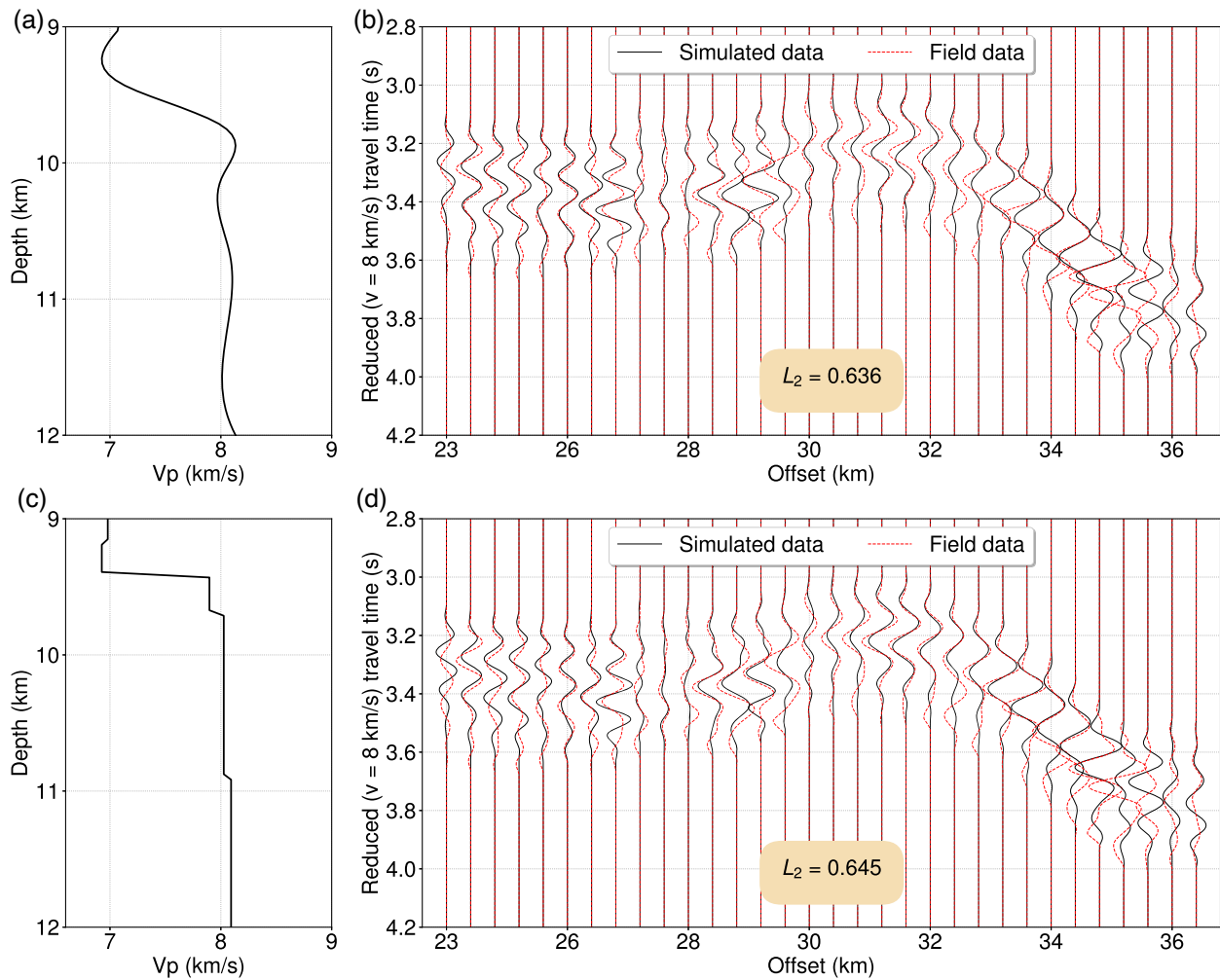


Figure 26. The maximum a posteriori probability (MAP) models for (a) the cubic spline interpolation and (c) the hybrid model parametrization. (b) and (d) show the comparison between the field OBS data and the data simulated using MAP for the two model parametrizations. The offset increment for the plotted traces is about 0.4 km.

searches the model space directly, and estimates the posterior distribution of the seismic velocity model. Compared with common MCMC methods (e.g. the Metropolis–Hastings algorithm) which require the number of unknown (e.g. the number of velocity layers for 1-D inversion) as a fixed prior, the proposed method is able to explore model space with a variable number of parameters, making the inversion workflow completely data-driven and simplifying the inversion preparation work.

To better simulate the physics of seismic wave propagation, 2-D wave equation modelling is performed to take into account the lateral heterogeneity in the upper crust, as observed in the velocity model from traveltime tomography. The computational time per MCMC chain using 10 CPU cores is about 50 hr for 100 000 iterations in the synthetic studies, and about 70 hr for the field study using one OBS gather. MCMC sampling usually requires a large number of iterations to find the global optimum, and the required iteration increases dramatically with the increasing number of parameters (the curse of dimensionality; Sambridge *et al.* 2006). Limited by computational resources, we use a sparse model parametrization with a 1-D assumption for velocity models close to Moho, for the purpose of reducing model dimension and consequently an efficient search of the model space. The 1-D assumption can be reasonable for Moho structure inversion within the offset range for

one common-OBS seismic gather, but may be not able to describe the model complexity for large-scale seismic survey or when there are complex geological settings. For future studies, it will be important to use a more advanced sparse model parametrization, for example, wavelet parametrization (Hawkins & Sambridge 2015), to allow a balance between the ability for characterizing complex (at least 2-D) earth models and minimizing the number of unknowns. Gradient information from adjoint-state FWI can be used for accelerating the model space exploration of Monte Carlo search, for example in Hamiltonian Monte Carlo (Biswas & Sen 2017; Fichtner *et al.* 2018; Fichtner & Zunino 2019; Zhao *et al.* 2019; Gebraad *et al.* 2020), therefore may allow us to use high-dimensional model parametrization. Variational inference may also provide an efficient solution for Bayesian inference problems (Zhang & Curtis 2020).

In the present study, the model obtained from traveltime tomography is preserved down to 9 km depth. It is worthwhile to mention that the resolution of traveltime tomography is significantly lower than that of the FWI, and therefore there may be potential travel-time errors when simulating seismograms using the tomographic model. The possible errors in the tomographic model may shift the Moho transition towards shallower (overestimated crust velocity) or deeper (underestimated crust velocity) depths. It would be useful to perform adjoint-state FWI for improving the oceanic crust

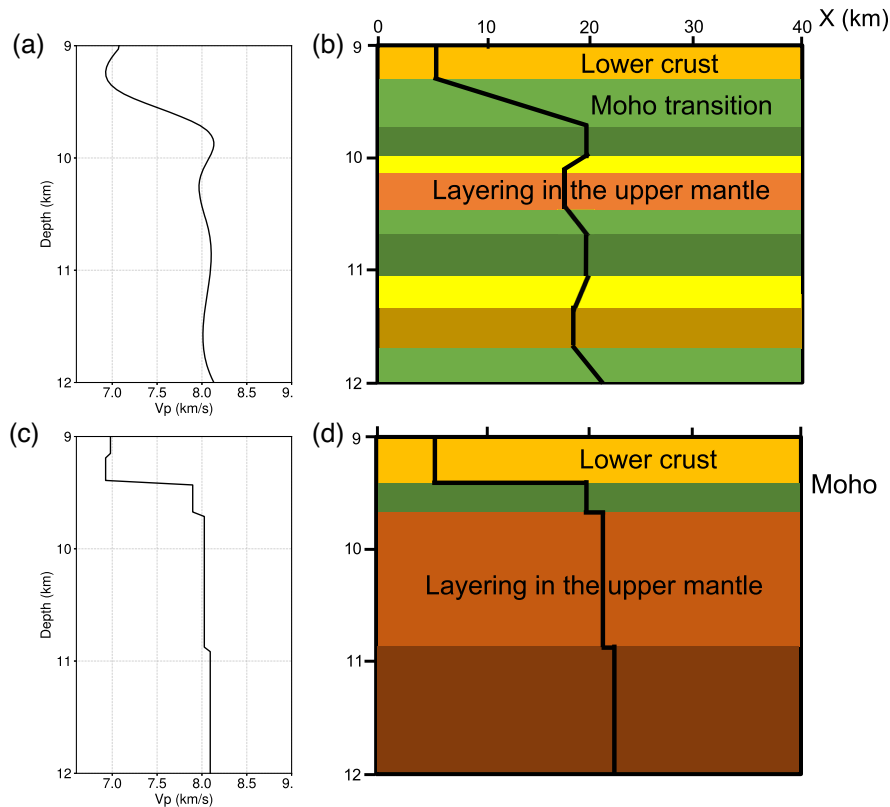


Figure 27. (a) and (c) are the MAP models from Figs 26(a) and (c); (b) and (d) are the schematic diagrams showing the Moho and the layering structures in the upper mantle from the inferred velocity models.

model for better accuracy, or conduct a complete Bayesian inversion workflow for the whole model including both crust and Moho (which requires more efficient sampling algorithms). With these in mind, this study is more interested in characterizing the Moho structure than its absolute depth. This is left for future studies.

The acoustic assumption is an approximation to the (visco)elastic earth, and has been widely used for FWI (Tarantola 1984; Virieux & Operto 2009; Kamei *et al.* 2013). With many successful application for acoustic FWI, we should bear in mind that many of them only use the turning waves (Kamei *et al.* 2013). Górszczyk *et al.* (2017) presented a robust workflow for OBS FWI in a complex subduction zone, with data constraints including not only the more linearized events of refracted and diving waves, but also wide angle Moho reflections. However, a very accurate velocity model is required in the vicinity of Moho in order to avoid getting trapped into local minima. In this study, we use mainly wide-aperture PmP reflections around critical angles as data constraints for the Moho structure. Previous studies (Chapman *et al.* 2014; Guo *et al.* 2019) indicate that there can be large waveform difference in both amplitude and phase, between data predicted by acoustic and elastic wave equations, especially around critical angles. Fig. 28 shows a comparison of elastic and acoustic wave modelling using the velocity profile in Fig. 2(a) and the acquisition geometry in Fig. 3. The *S*-wave velocity is derived from the *P*-wave velocity using an empirical relation following Brocher (2005), with velocities in the sea water setting to zero. The green oval indicates the Moho reflection (PmP). The waveform matches well for the first arrival (Pg and Pn) for acoustic and elastic modelling, but there is more difference for PmP. Moho reflection from elastic modelling is generally weaker than that from acoustic modelling, because of the energy conversion between *P* and

S waves. An elastic wave equation modelling is expected to provide a more realistic modelling of the physics for wave propagation in the solid earth, and therefore have the potential for improving the field data inversion results, at the trade-off of a significant increase on the computational cost. This is left for future studies.

6 CONCLUSIONS

In the presence of a high-velocity contrast, such as at Moho, FWI using wide-aperture seismic data is a strong nonlinear problem because of the abrupt changes in reflection coefficients around the critical angles and interferences from different seismic phases, therefore the widely used linearized inversion approach may not work. We propose to use the nonlinear trans-dimensional MCMC method for solving the seismic FWI problem in the context of Moho structure building, with the state-of-the-art seismic waveform modelling by finite-difference solution of wave equation. The use of 2-D wave equation modelling allows us to incorporate the wave physics for the heterogeneous upper crust structure, and the 1-D model parametrization near Moho enables an efficient search of the trans-dimensional model space. Trans-dimensional MCMC provides a complete data-driven probabilistic solution for the nonlinear FWI problem, with the outcome including the posterior distribution for both the velocity–depth structure, and the model complexity (the optimal number of parameters for describing the model).

With very weak prior information, synthetic studies using four different Moho scenarios demonstrate the robustness of the trans-dimensional MCMC method, with high probabilities in the posterior distribution consistent with the target velocity structure. Inversion results from the field OBS data in the Mid-Atlantic Ocean indicates

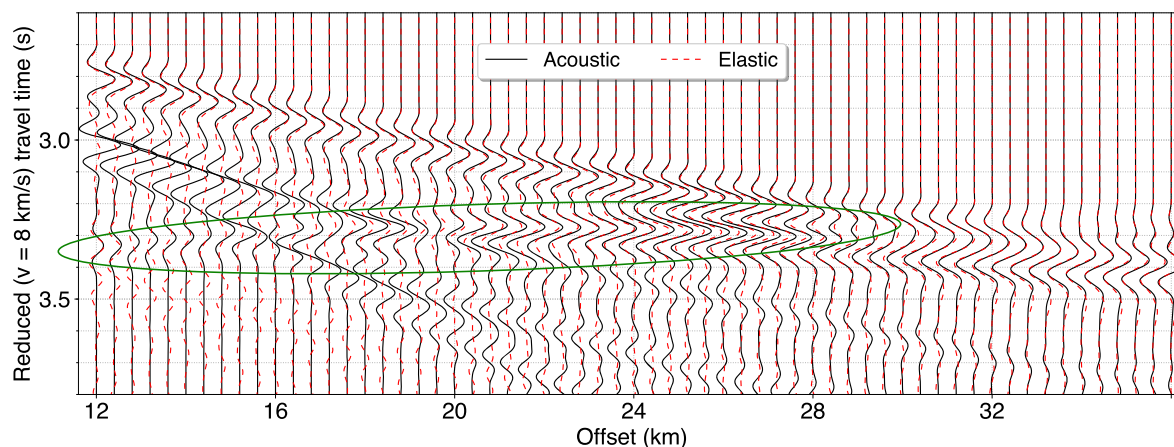


Figure 28. A comparison of simulated seismograms using acoustic (solid black line) and elastic (dashed red line) wave equations. The acquisition geometry is the same with that in Fig. 3. We use the P -wave velocity profile in Fig. 2(a); the S -wave velocity is derived from the P -wave velocity using an empirical relation following Brocher (2005), with velocities in the sea water setting to zero. The green oval indicates the Moho reflection (PmP).

that Moho in the study area can be a transition zone of around 0.7 km, or it can also be a sharp boundary with velocities from around 7 km s^{-1} in the lower crust to 8 km s^{-1} of the upper mantle; both provide similar waveform match for the field data. The ambiguity comes from the resolution limit of the band-limited seismic data and from the limited scattering angle range for observed Moho reflection. The inferred posterior distribution of interfaces (velocity nodes) strongly coincides with the depth of velocity discontinuity. Preliminary results from our study indicate layered structure in the vicinity of Moho.

ACKNOWLEDGEMENTS

The OBS data were acquired during the LITHOS experiment on board the German R/V Maria S. Merian in 2017, in the leadership of Dr Ingo Grevemeyer. This research was funded by the Deep Earth Imaging Future Science Platform, CSIRO. This project was funded by the European Research Council Advanced Grant agreement No. 339442 TransAtlanticILAB. The cruise with the ID MSM69 was funded by the German Science Foundation (DFG). This work was supported by resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia. We also thank the computational resources supported by S-CAPAD. Finally, we thank Editor Dr Michal Malinowski, and reviewer Dr Andrzej Górszczyk and an anonymous reviewer for their constructive comments which significantly improved the paper.

REFERENCES

- Agostinetti, N.P. & Malinverno, A., 2010. Receiver function inversion by trans-dimensional Monte Carlo sampling, *Geophys. J. Int.*, **181**(2), 858–872.
- Aki, K. & Richards, P.G., 1980. *Quantative Seismology: Theory and Methods*, Elsevier.
- Auld, B.A., 1973. *Acoustic Fields and Waves in Solids: Volume I*, John Wiley Sons.
- Beller, S., Monteiller, V., Operto, S., Nolet, G., Paul, A. & Zhao, L., 2018. Lithospheric architecture of the south-western Alps revealed by multi-parameter teleseismic full-waveform inversion, *Geophys. J. Int.*, **212**(2), 1369–1388.
- Biswas, R. & Sen, M., 2017. 2D full-waveform inversion and uncertainty estimation using the reversible jump Hamiltonian Monte Carlo, in 2017 SEG International Exposition and Annual Meeting, Society of Exploration Geophysicists.
- Bodin, T. & Sambridge, M., 2009. Seismic tomography with the reversible jump algorithm, *Geophys. J. Int.*, **178**(3), 1411–1436.
- Bodin, T., Sambridge, M., Tkalčić, H., Arroucau, P., Gallagher, K. & Rawlinson, N., 2012. Transdimensional inversion of receiver functions and surface wave dispersion, *J. geophys. Res.*, **117**(B2), doi:10.1029/2011JB008560.
- Bozdağ, E., Trampert, J. & Tromp, J., 2011. Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements, *Geophys. J. Int.*, **185**(2), 845–870.
- Brocher, T.M., 2005. Empirical relations between elastic wavespeeds and density in the earth's crust, *Bull. seism. Soc. Am.*, **95**(6), 2081–2092.
- Brocher, T.M., Karson, J.A. & Collins, J.A., 1985. Seismic stratigraphy of the oceanic Moho based on ophiolite models, *Geology*, **13**(1), 62–65.
- Buehler, J. & Shearer, P., 2017. Uppermost mantle seismic velocity structure beneath USArray, *J. geophys. Res.*, **122**(1), 436–448.
- Burdick, S. & Lekić, V., 2017. Velocity variations and uncertainty from transdimensional P -wave tomography of North America, *Geophys. J. Int.*, **209**(2), 1337–1351.
- Cary, P. & Chapman, C., 1988. Automatic 1-D waveform inversion of marine seismic refraction data, *Geophys. J. Int.*, **93**(3), 527–546.
- Chapman, C. & Orcutt, J., 1985. The computation of body wave synthetic seismograms in laterally homogeneous media, *Rev. Geophys.*, **23**(2), 105–163.
- Chapman, C.H., 1978. A new method for computing synthetic seismograms, *Geophys. J. Int.*, **54**(3), 481–518.
- Chapman, C.H., Hobro, J.W. & Robertsson, J.O., 2014. Correcting an acoustic wavefield for elastic effects, *Geophys. J. Int.*, **197**(2), 1196–1214.
- Dadi, S., Gibson, R.L. & Sager, W.W., 2018. Understanding the late-stage evolution of Shatsky Rise using transdimensional acoustic impedance inversion, *J. geophys. Res.*, **123**(2), 1576–1590.
- Dettmer, J., Dosso, S.E. & Holland, C.W., 2010. Trans-dimensional geoaoustic inversion, *J. acoust. Soc. Am.*, **128**(6), 3393–3405.
- Dettmer, J., Hawkins, R., Cummins, P.R., Hossen, J., Sambridge, M., Hino, R. & Inazu, D., 2016. Tsunami source uncertainty estimation: the 2011 Japan tsunami, *J. geophys. Res.*, **121**(6), 4483–4505.
- Fichtner, A. & Zunino, A., 2019. Hamiltonian nullspace shuttles, *Geophys. Res. Lett.*, **46**(2), 644–651.
- Fichtner, A., Bunge, H.-P. & Igel, H., 2006. The adjoint method in seismology: I. Theory, *Phys. Earth planet. Inter.*, **157**(1–2), 86–104.
- Fichtner, A., Zunino, A. & Gebrad, L., 2018. Hamiltonian Monte Carlo solution of tomographic inverse problems, *Geophys. J. Int.*, **216**(2), 1344–1363.

- Forbriger, T., Groos, L. & Schäfer, M., 2014. Line-source simulation for shallow-seismic data. Part 1: theoretical background, *Geophys. J. Int.*, **198**(3), 1387–1404.
- Galetti, E., Curtis, A., Baptie, B., Jenkins, D. & Nicolson, H., 2017. Transdimensional Love-wave tomography of the British Isles and shear-velocity structure of the East Irish Sea Basin from ambient-noise interferometry, *Geophys. J. Int.*, **208**(1), 36–58.
- Ganley, D., 1981. A method for calculating synthetic seismograms which include the effects of absorption and dispersion, *Geophysics*, **46**(8), 1100–1107.
- Gebräad, L., Boehm, C. & Fichtner, A., 2020. Bayesian elastic full-waveform inversion using Hamiltonian Monte Carlo, *J. geophys. Res.*, **125**(3), e2019JB018428, doi:10.1029/2019JB018428.
- Górszczyk, A., Operto, S. & Malinowski, M., 2017. Toward a robust workflow for deep crustal imaging by FWI of OBS data: the eastern Nankai Trough revisited, *J. geophys. Res.*, **122**(6), 4601–4630.
- Gouveia, W.P. & Scales, J.A., 1998. Bayesian seismic waveform inversion: parameter estimation and uncertainty analysis, *J. geophys. Res.*, **103**(B2), 2759–2779.
- Grad, M., Tiira, T. & Group, E.W., 2009. The Moho depth map of the European plate, *Geophys. J. Int.*, **176**(1), 279–292.
- Green, P.J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, **82**(4), 711–732.
- Green, P.J., 2003. *Trans-dimensional Markov chain Monte Carlo*, Oxford University Press, 179–198.
- Guo, P., McMechan, G.A. & Ren, L., 2019. Modeling the viscoelastic effects in P-waves with modified viscoacoustic wave propagation, *Geophysics*, **84**(6), T381–T394.
- Guo, P., Visser, G. & Saygin, E., 2020. Bayesian trans-dimensional full waveform inversion: synthetic and field data application, *Geophys. J. Int.*, **222**(1), 610–627.
- Hawkins, R. & Sambridge, M., 2015. Geophysical imaging using trans-dimensional trees, *Geophys. J. Int.*, **203**(2), 972–1000.
- Hawkins, R., Bodin, T., Sambridge, M., Choblet, G. & Husson, L., 2019. Trans-dimensional surface reconstruction with different classes of parameterization, *Geochem. Geophys. Geosyst.*, **20**(1), 505–529.
- Hrubcová, P., Vavříčuk, V., Boušková, A. & Horálek, J., 2013. Moho depth determination from waveforms of microearthquakes in the West Bohemia/Vogtland swarm area, *J. geophys. Res.*, **118**(1), 120–137.
- Jian, H., Singh, S.C., Chen, Y.J. & Li, J., 2017. Evidence of an axial magma chamber beneath the ultraslow-spreading Southwest Indian Ridge, *Geology*, **45**(2), 143–146.
- Kamei, R., Pratt, R. & Tsuji, T., 2013. On acoustic waveform tomography of wide-angle obs data—strategies for pre-conditioning and inversion, *Geophys. J. Int.*, **194**(2), 1250–1280.
- Killingbeck, S., Livermore, P., Booth, A. & West, L., 2018. Multimodal layered transdimensional inversion of seismic dispersion curves with depth constraints, *Geochem. Geophys. Geosyst.*, **19**(12), 4957–4971.
- Malinverno, A., 2002. Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem, *Geophys. J. Int.*, **151**(3), 675–688.
- Martin, R. & Komatitsch, D., 2009. An unsplit convolutional perfectly matched layer technique improved at grazing incidence for viscoelastic wave equation, *Geophys. J. Int.*, **179**(1), 333–344.
- McGlashan, N., Brown, L. & Kay, S., 2008. Crustal thickness in the central Andes from teleseismically recorded depth phase precursors, *Geophys. J. Int.*, **175**(3), 1013–1022.
- Métivier, L., Brossier, R., Mèrigot, Q., Oudet, E. & Virieux, J., 2016. Measuring the misfit between seismograms using an optimal transport distance: application to full waveform inversion, *Geophys. J. Int.*, **205**(1), 345–377.
- Mosegaard, K. & Tarantola, A., 1995. Monte carlo sampling of solutions to inverse problems, *J. geophys. Res.*, **100**(B7), 12 431–12 447.
- Mosegaard, K., Singh, S., Snyder, D. & Wagner, H., 1997. Monte Carlo analysis of seismic reflections from Moho and the W reflector, *J. geophys. Res.*, **102**(B2), 2969–2981.
- Pica, A., Diet, J. & Tarantola, A., 1990. Nonlinear inversion of seismic reflection data in a laterally invariant medium, *Geophysics*, **55**(3), 284–292.
- Plessix, R.-E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophys. J. Int.*, **167**(2), 495–503.
- Pratt, R.G., Shin, C. & Hick, G., 1998. Gauss–Newton and full Newton methods in frequency–space seismic waveform inversion, *Geophys. J. Int.*, **133**(2), 341–362.
- Ray, A., Sekar, A., Hoversten, G.M. & Albertin, U., 2016. Frequency domain full waveform elastic inversion of marine seismic data from the Alba field using a Bayesian trans-dimensional algorithm, *Geophys. J. Int.*, **205**(2), 915–937.
- Ray, A., Kaplan, S., Washbourne, J. & Albertin, U., 2017. Low frequency full waveform seismic inversion within a tree based Bayesian framework, *Geophys. J. Int.*, **212**(1), 522–542.
- Sambridge, M., 1999. Geophysical inversion with a neighbourhood algorithm—II. Appraising the ensemble, *Geophys. J. Int.*, **138**(3), 727–746.
- Sambridge, M. & Drijkoningen, G., 1992. Genetic algorithms in seismic waveform inversion, *Geophys. J. Int.*, **109**(2), 323–342.
- Sambridge, M. & Mosegaard, K., 2002. Monte Carlo methods in geophysical inverse problems, *Rev. Geophys.*, **40**(3), 3–1.
- Sambridge, M., Gallagher, K., Jackson, A. & Rickwood, P., 2006. Trans-dimensional inverse problems, model comparison and the evidence, *Geophys. J. Int.*, **167**(2), 528–542.
- Sambridge, M., Bodin, T., Gallagher, K. & Tkalčić, H., 2013. Transdimensional inference in the geosciences, *Phil. Trans. R. Soc. A*, **371**(1984), 20110547.
- Saygin, E., et al., 2015. Imaging architecture of the Jakarta Basin, Indonesia with transdimensional inversion of seismic noise, *Geophys. J. Int.*, **204**(2), 918–931.
- Sen, M.K. & Biswas, R., 2017. Transdimensional seismic inversion using the reversible jump Hamiltonian Monte Carlo algorithm, *Geophysics*, **82**(3), R119–R134.
- Shaw, P.R., 1986. The accuracy of models derived by WKBJ waveform inversion, *Geophys. J. Int.*, **85**(2), 291–313.
- Shipp, R.M. & Singh, S.C., 2002. Two-dimensional full wavefield inversion of wide-aperture marine seismic streamer data, *Geophys. J. Int.*, **151**(2), 325–344.
- Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, **49**(8), 1259–1266.
- Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.*, **160**(1), 195–216.
- Vaddineni, V.A., Singh, S.C. & Grevemeyer, I., 2018. Tomographic study of oceanic lithosphere from 0 to 25 Ma in the equatorial Atlantic Ocean using wide angle OBS data, in *American Geophysical Union, Fall Meeting*, abstract T23E–0410.
- Van Avendonk, H.J., Harding, A.J., Orcutt, J.A. & McClain, J.S., 1998. A two-dimensional tomographic study of the Clipperton transform fault, *J. geophys. Res.*, **103**(B8), 17885–17899.
- Van Avendonk, H.J., Shillington, D.J., Holbrook, W.S. & Hornbach, M.J., 2004. Inferring crustal structure in the Aleutian island arc from a sparse wide-angle seismic data set, *Geochem. Geophys. Geosyst.*, **5**(8), doi:10.1029/2003GC000664.
- Vigh, D., Jiao, K., Watts, D. & Sun, D., 2014. Elastic full-waveform inversion application using multicomponent measurements of seismic data collection, *Geophysics*, **79**(2), R63–R77.
- Virieux, J., 1986. P-SV wave propagation in heterogeneous media: velocity-stress finite-difference method, *Geophysics*, **51**(4), 889–901.
- Virieux, J. & Operto, S., 2009. An overview of full-waveform inversion in exploration geophysics, *Geophysics*, **74**(6), WCC1–WCC26.
- Visser, G., Guo, P. & Saygin, E., 2019. Bayesian transdimensional seismic full waveform inversion with a dipping layer parameterization, *Geophysics*, **84**(6), R845–R858.

- Zelt, C.A., Sain, K., Naumenko, J.V. & Sawyer, D.S., 2003. Assessment of crustal velocity models using seismic refraction and reflection tomography, *Geophys. J. Int.*, **153**(3), 609–626.
- Zhang, X. & Curtis, A., 2020. Seismic tomography using variational inference methods, *J. geophys. Res.*, **125**(4), e2019JB018589, doi:10.1029/2019JB018589.
- Zhao, Z., Sen, M.K., *et al.*, 2019. A gradient based MCMC method for FWI and uncertainty analysis, in *SEG International Exposition and Annual Meeting*, Society of Exploration Geophysicists.
- Zhu, D. & Gibson, R., 2018. Seismic inversion and uncertainty quantification using transdimensional Markov chain Monte Carlo method, *Geophysics*, **83**(4), 1–76.